

Bounding stationary averages of polynomial diffusions via semidefinite programming

Juan Kuntz* Michela Ottobre† Guy-Bart Stan‡ Mauricio Barahona§

June 6, 2016

Abstract

We introduce an algorithm based on semidefinite programming that yields increasing (resp. decreasing) sequences of lower (resp. upper) bounds on polynomial stationary averages of diffusions with polynomial drift vector and diffusion coefficients. The bounds are obtained by optimising an objective, determined by the stationary average of interest, over the set of real vectors defined by certain linear equalities and semidefinite inequalities which are satisfied by the moments of any stationary measure of the diffusion. We exemplify the use of the approach through several applications: a Bayesian inference problem; the computation of Lyapunov exponents of linear ordinary differential equations perturbed by multiplicative white noise; and a reliability problem from structural mechanics. Additionally, we prove that the bounds converge to the infimum and supremum of the set of stationary averages for certain SDEs associated with the computation of the Lyapunov exponents, and we provide numerical evidence of convergence in more general settings.

1 Introduction. Stochastic differential equations (SDEs) and the diffusion processes they generate are important modelling tools in numerous scientific fields, such as chemistry, economics, physics, biology, finance, and epidemiology [22]. Stationary measures are to SDEs what fixed points are to deterministic systems: if the SDE is stable, then its stationary measures determine its long-term statistics. More concretely, both the ensemble averages and the time averages of the process converge to averages with respect to a stationary measure (which we call stationary averages). For the majority of SDEs, there are no known analytical expressions for their stationary measures. Consequently, large efforts have been directed at developing computational tools that estimate stationary averages and, more generally, at developing tools that can be used to study the long-term behaviour of SDEs. Among these, most prominent are Monte Carlo discretisation schemes, PDE methods (finite-difference, finite-element, Galerkin methods), path integral methods, moment closure methods, and linearisation techniques (see, e.g., [39, 6, 43, 45]).

The purpose of this paper is to introduce a new algorithm that provides an alternative approach to the analysis of stationary measures. It uses semidefinite programming to compute hard bounds on polynomial stationary averages of polynomial diffusions. Our approach has distinct advantages: (i) it returns monotone sequences of both upper and lower bounds on the stationary averages,

*Department of Mathematics and Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (corresponding author: jk208@ic.ac.uk).

†Mathematics Department, Heriot Watt University, Edinburgh, EH14 4AS, UK (m.ottobre@hw.ac.uk).

‡Department of Bioengineering, Imperial College London, London, SW7 2AZ, UK (g.stan@ic.ac.uk).

§Department of Mathematics, Imperial College London, London, SW7 2AZ, UK (corresponding author: m.barahona@ic.ac.uk).

hence quantifying precisely the uncertainty in our knowledge of the stationary averages; (ii) no assumptions are required on the uniqueness of the stationary measures of the SDE; and (iii) the availability of high quality SDP solvers and of the modelling package GloptiPoly 3 drastically reduces the investment of effort and specialised knowledge required to implement the algorithm for the analysis of a given diffusion process.

1.1 Problem definition. We consider \mathbb{R}^n -valued diffusion processes $X := \{X_t : t \geq 0\}$ satisfying stochastic differential equations of the form

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad X_0 = Z, \quad (1.1)$$

where the entries of the drift vector $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and the diffusion coefficients $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ are *polynomials*. In the above, $W := \{W_t : t \geq 0\}$ is a standard \mathbb{R}^m -valued Brownian motion and the initial condition Z is a Borel measurable random variable. We assume that the underlying filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ satisfies the usual conditions.

A Borel probability measure π on \mathbb{R}^n is a *stationary* (or *invariant*) *measure* of the dynamics (1.1) if

$$Z \sim \pi \quad \Rightarrow \quad X_t \sim \pi \text{ for all } t \geq 0,$$

where we use the notation $Y \sim \pi$ to mean that the random variable Y has law π . The set of stationary measures of (1.1) is denoted \mathcal{P} .

The problem we address here is how to estimate stationary averages of the form

$$\pi(f) := \int f(x) \pi(dx) \quad (1.2)$$

in a systematic, computationally efficient manner. We present an algorithm that yields bounds on averages (1.2) when f is a *polynomial*. Hence, our algorithm can be used to bound the *moments* of the stationary measures of (1.1).

More precisely, the algorithm returns lower and upper bounds on the set

$$B_{f,\mathcal{G}}^d := \{\pi(f) : \pi \in \mathcal{P} \text{ has finite } d\text{-th order moments and support contained in } \mathcal{G}\}, \quad (1.3)$$

where \mathcal{G} is a given real algebraic variety

$$\mathcal{G} := \{x \in \mathbb{R}^n : g_1(x) = 0, \dots, g_\ell(x) = 0\}, \quad (1.4)$$

for given polynomials g_1, \dots, g_ℓ . The case $\mathcal{G} = \mathbb{R}^n$ corresponds to $\ell = 0$.

Stationary averages of the form (1.2) and the set (1.3) are of broad interest: if X enjoys some mild stability and regularity properties [32], the stationary averages (1.2) provide succinct, quantitative information about the *long-term* behaviour of X . In particular, for almost every sample path $t \mapsto X_t(\omega)$ (that is, \mathbb{P} -almost every $\omega \in \Omega$), there exists a $\pi \in \mathcal{P}$ such that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X_s(\omega)) ds = \int f(x) \pi(dx), \quad (1.5)$$

see Theorem 2.2 for a formal statement. Furthermore, for appropriately chosen d , the set $B_{f,\mathcal{G}}^d$ is the set of limits (1.5) where $t \mapsto X_t(\omega)$ is any sample path contained in \mathcal{G} :

$$B_{f,\mathcal{G}}^d = \left\{ \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X_s(\omega)) ds : \omega \in \Omega_{\mathcal{G}} \right\}, \quad (1.6)$$

where $\Omega_{\mathcal{G}}$ is the subset of samples $\omega \in \Omega$ such that $X_t(\omega)$ belongs to \mathcal{G} for all $t \geq 0$, see [32, §8]. Similar considerations also apply to the ensemble averages $\mathbb{E}[f(X_t)]$ (see the remark after Theorem 2.2). Therefore, bounds on the set $B_{f,\mathcal{G}}^d$ equip us with quantitative information on the long-term behaviour of the paths of X that are contained in \mathcal{G} .

Remark (Linking the long-term behaviour to the stationary measures): Although our work is motivated by the study of the long-term behaviour of a diffusion X , the scope of this paper is restricted to the problem of bounding the stationary averages (1.2) and the associated set (1.3). It is important to remark that to connect the long-term behaviour of X with the set $B_{f,\mathcal{G}}^d$ (and the bounds our algorithm returns) it is necessary to establish separately:

- (a) the existence of stationary measures of (1.1) with support contained in \mathcal{G} and the finiteness of their moments up to order d ;
- (b) the convergence of the time averages, i.e. verify that the limit (1.5) holds for the diffusion at hand;
- (c) that, for the initial conditions of interest, X takes values in \mathcal{G} .

A straightforward way to verify (c) is to apply Itô's formula and check that

$$dg_1(X_t) = dg_2(X_t) = \dots = dg_\ell(X_t) = 0.$$

If the initial condition Z takes values in \mathcal{G} , then X clearly takes values in \mathcal{G} as well. Establishing (a) and (b) typically requires additional proofs beyond the scope of this paper, and has been studied extensively elsewhere (see [32, 33, 20]). We point out that whether or not conditions (a)-(c) hold, the algorithm still yields bounds on $B_{f,\mathcal{G}}^d$. Without proving (a)-(c), it may simply be that the set (1.3) is empty, or that the relationships (1.5) or (1.6) do not exist. To make the paper self-contained, we recall briefly in Section 2.2 some simple conditions that we use in our later examples to establish (a) and (b).

1.2 Brief description of algorithm. Mathematically, our approach consists of four steps:

- (1) We derive a finite set of linear equalities satisfied by the moments of any stationary measure of (1.1) (see Lemma 2.3 and Section 3.1). Such a system of equalities is often underdetermined (see Example 3.3) and therefore admits infinitely many solutions.
- (2) To rule out spurious solutions, we exploit the fact that the solutions must be the moments of a probability measure, hence must satisfy extra constraints (e.g. even moments cannot be negative). We use well known results [25] to construct semidefinite inequalities (so called *moment constraints*) that are satisfied by moments of any probability measure with support contained in \mathcal{G} . This is done in Section 3.2.
- (3) Exploiting the fact that f is a polynomial, we rewrite the stationary average $\pi(f)$ as a linear combination of the moments of π .
- (4) By maximising (resp. minimising) the linear combination over the set of all vectors of real numbers that satisfy *both* the linear equalities and the semidefinite inequalities, we obtain an upper (resp. lower) bound on the set $B_{f,\mathcal{G}}^d$.

Computationally, to find the upper (or lower) bound we solve a semidefinite program, a particularly tractable class of convex optimisation problems for which high-quality solvers are freely available online. The semidefinite constraints in (2), popularised by Lasserre and co-authors (see [25] and references therein), can be implemented via the freely-available, user-friendly package Glop-tiPoly 3 [13], which makes the approach described here accessible to non-experts.

Remark (Convergence of the algorithm): Concerning moment approaches, a question that often arises is that of convergence of the algorithm. In the context of this paper, this question takes the following form. Suppose that we want to obtain bounds on the set

$$B_{f,\mathcal{G}}^\infty := \bigcap_{d=d_f}^\infty B_{f,\mathcal{G}}^d,$$

where d_f is the degree of f . Note that if \mathcal{G} is compact, then any measure with support on \mathcal{G} has all moments finite, thus $B_{f,\mathcal{G}}^{d_f} = B_{f,\mathcal{G}}^{d_f+1} = \dots = B_{f,\mathcal{G}}^\infty$; hence $B_{f,\mathcal{G}}^\infty$ is the only set of interest. As will become clear later, repeated applications of our algorithm yield both an increasing sequence of lower bounds and a decreasing sequence of upper bounds on $B_{f,\mathcal{G}}^\infty$. The algorithm is said to converge if these sequences converge to the infimum and supremum, respectively, of $B_{f,\mathcal{G}}^\infty$. In general, the algorithm presented in this paper is not guaranteed to converge in the above sense. However, our numerics indicate that the algorithm often converges in practice (see Examples 3.4, 4.1 and 4.3). In Section 4.2, we do prove convergence for SDEs related to the Lyapunov exponents of linear ordinary differential equations (ODEs) perturbed by multiplicative white noise. The question of convergence is of theoretical interest, but regardless of its answer, the bounds computed are still valid and are often still appropriate for the application in question (e.g., see the examples in Section 4).

1.3 Related literature and contributions of the paper. Computational methods that yield bounds on functionals of Markov processes by combining linear equalities (arising from the definitions of the functional and the process) and moment constraints have appeared in the literature. We refer to this class of methods as *generalised moment approaches* (GMAs). The various GMAs differ in the type of Markov processes and moment constraints they consider. The ideas underlying GMAs were first discussed in [4, 5] where they were used to obtain analytical bounds on moments for measure-valued diffusions. The first GMA was presented in E. Schwerer’s PhD thesis [40] in the context of jump processes with bounded generators and reflected Brownian motion on the non-negative orthant. In [15, §12.4], the authors present a GMA that yields bounds on the moments of stationary measures of discrete-time Feller Markov chains. In [12], analogous techniques are used to bound the moments of the stationary measures of diffusion approximations of the Wright-Fisher model on the unit simplex. GMAs have also been proposed to solve optimal control problems [14, 11], to estimate exit times [10, 26], and to price financial derivatives [27, 7].

The contributions of this paper are as follows. Whereas [40, 12, 11] consider only stationary averages of specific SDEs, we introduce GMAs to the setting of general polynomial diffusions over unbounded domains—this requires setting up the technical background of Lemma 2.3. Also in contrast with [40, 12, 11], we employ moment constraints that lead to SDPs instead of linear programs. In Section 4.2, we prove convergence of our algorithm for SDEs related to the Lyapunov exponents of linear ODEs perturbed by multiplicative white noise. To the best of our knowledge, this is the first such result in the setting of stationary averages of diffusion processes. The remaining contributions are the applications of the algorithm to several examples of interest. Section 4.1 explains how the algorithm can be combined with the ideas underlying the Metropolis Adjusted Langevin Algorithm (a Markov chain Monte Carlo algorithm) to carry out numerical integration with respect to certain target measures, and then applies it to a simple Bayesian inference problem. In Section 4.2, we use our algorithm to obtain bounds on the Lyapunov exponents of linear differential equations perturbed by multiplicative white noise, and we show that in this case our approach is both sufficient and necessary (i.e., with enough computation power, it yields lower (upper) bounds arbitrarily close to the minimum (maximum) Lyapunov exponent, see Theorem 4.2). Finally, in Section 4.3 we explain how the algorithm can be extended to yield bounds on stationary averages

$\pi(f)$ where f is piecewise polynomial, and we use this extension to tackle a reliability problem from structural mechanics.

2 Background and preliminaries.

2.1 Notation Throughout this paper we use the following notation:

- $\mathbb{E}[\cdot]$ denotes expectation with respect to the underlying probability measure \mathbb{P} , and we use X_t and $X(t)$ interchangeably.
- Given a function $h : \mathbb{R}^n \rightarrow \mathbb{R}$, $\partial_i h$ denotes its partial derivative with respect to its i^{th} argument; $\partial_{ij} h := \partial_i \partial_j h$; ∇h denotes its gradient vector; and $\nabla^2 h$ denotes its Hessian matrix.
- Suppose that \mathcal{M} is a smooth manifold. $C^2(\mathcal{M})$ denotes the set of real-valued, twice continuously differentiable functions on \mathcal{M} .
- For any two matrices $A, B \in \mathbb{R}^{n \times m}$,

$$\langle A, B \rangle := \sum_{i=1}^n \sum_{j=1}^m A_{ij} B_{ij}$$

denotes the trace inner product of A and B , and $\|A\| := \sqrt{\langle A, A \rangle}$ denotes the Frobenius norm of A .

- Let \mathbb{N}^n be the set of n -tuples $\alpha := (\alpha_1, \dots, \alpha_n)$ of natural numbers $\alpha_i \in \mathbb{N}$. Let \mathbb{N}_d^n be the subset of n -tuples such that $|\alpha| := \alpha_1 + \dots + \alpha_n \leq d$. The cardinality of \mathbb{N}_d^n is $r(d) := \binom{n+d}{d}$. We define the sum of two tuples $\alpha, \beta \in \mathbb{N}_d^n$ to be the tuple $\alpha + \beta := (\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n)$.
- The space of real-valued vectors indexed by \mathbb{N}_d^n , $\{y : y_\alpha \in \mathbb{R}, \alpha \in \mathbb{N}_d^n\}$, is isomorphic to $\mathbb{R}^{r(d)}$ and we make no distinction between them. Similarly for the space of real-valued matrices indexed by \mathbb{N}_d^n , $\{M : M_{\alpha\beta} \in \mathbb{R}, \alpha, \beta \in \mathbb{N}_d^n\}$, and $\mathbb{R}^{r(d) \times r(d)}$. With this in mind, we denote the standard inner product on $\mathbb{R}^{r(d)}$ as

$$\langle y, z \rangle := \sum_{\alpha \in \mathbb{N}_d^n} y_\alpha z_\alpha, \quad y, z \in \mathbb{R}^{r(d)},$$

and the standard outer product on $\mathbb{R}^{r(d)}$ as

$$(y \otimes z)w := \langle z, w \rangle y, \quad y, z, w \in \mathbb{R}^{r(d)}.$$

- Let $\alpha \in \mathbb{N}_d^n$ and $x \in \mathbb{R}^n$. Monomials are denoted as $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$, and $m_d(x) : \mathbb{R}^n \rightarrow \mathbb{R}^{r(d)}$ denotes the vector of monomials of degree d or less. Hence the α -th component of the $r(d)$ -dimensional vector $m_d(x)$ is given by

$$(m_d(x))_\alpha := x^\alpha.$$

Let $\mathbb{R}[x]_d$ denote the vector space of real polynomials on \mathbb{R}^n of degree at most d . The set $\{x^\alpha : \alpha \in \mathbb{N}_d^n\}$ is a basis (known as the canonical or monomial basis) of $\mathbb{R}[x]_d$, and so we can write any polynomial $p \in \mathbb{R}[x]_d$ as

$$p(x) = \sum_{\alpha \in \mathbb{N}_d^n} \mathbf{p}_\alpha x^\alpha = \langle \mathbf{p}, m_d(x) \rangle, \quad \forall x \in \mathbb{R}^n,$$

where $\mathbf{p} := (\mathbf{p}_\alpha)_{\alpha \in \mathbb{N}_d^n}$ in $\mathbb{R}^{r(d)}$ is the *vector of coefficients* of p . We denote the degree of any polynomial p with d_p .

- Let μ be a Borel measure on \mathbb{R}^n . A vector y in $\mathbb{R}^{r(d)}$ is the *vector of moments of order up to d* of μ if, for any $\alpha \in \mathbb{N}_d^n$, the α -th component of y is given by

$$y_\alpha = \mu(x^\alpha) := \int_{\mathbb{R}^n} x^\alpha \mu(dx),$$

assuming that the integrals are well defined.

2.2 Stability and regularity properties of X . We now briefly review well known properties of X and a relevant drift condition used in the examples below. Throughout this section we assume that \mathcal{G} is an $(n - \ell)$ -dimensional smooth submanifold of \mathbb{R}^n . For this to be the case, it is sufficient that the vectors $\nabla g_1(x), \nabla g_2(x), \dots, \nabla g_\ell(x)$ form a linearly independent set, for each x in \mathcal{G} , where the g_i 's are defined in (1.4).

Firstly, smoothness of the components of b and σ imply that (1.1) has a unique strong solution X , which is defined up to a stopping time τ_∞ [47]. The *generator* (or Kolmogorov operator) \mathcal{A} associated with (1.1) is the second order differential operator

$$\mathcal{A}h(x) := \langle \nabla h(x), b(x) \rangle + \frac{1}{2} \langle \nabla^2 h(x), a(x) \rangle = \sum_{i=1}^n b_i(x) \partial_i h(x) + \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \partial_{ij} h(x),$$

for $h \in C^2(\mathcal{G})$, $x \in \mathcal{G}$, and where $a := \sigma \sigma^T$ denotes the diffusion matrix of (1.1). It is well known that if \mathcal{A} is a hypoelliptic operator on $C^2(\mathcal{G})$ (see [16]), then (1.1) generates a strong Feller Markov process. This is the regularity property of X we use in our examples below. Although this condition can be replaced with weaker ones (e.g., X being a T-process [32]), such alternative conditions usually require more work to establish in practice. The stability properties we require are summarised as follows:

Condition 2.1 ([32, 33]). *Suppose that the paths of the diffusion X are contained in \mathcal{G} (that is, $\mathbb{P}(\{X_t \in \mathcal{G}, \forall 0 \leq t \leq \tau_\infty\}) = 1$), and that either one of the following conditions holds:*

(i) *the manifold \mathcal{G} is compact.*

(ii) *(Drift condition) there exists a function $u \in C^2(\mathcal{G})$ and a constant $c > 0$ such that for each $q \in \mathbb{R}$ the sub-level set*

$$\{x \in \mathcal{G} : u(x) \leq q\}$$

is compact and such that

$$\mathcal{A}u(x) \leq -cu(x),$$

holds for all x in \mathcal{G} except those in a compact subset of \mathcal{G} .

Theorem 2.2 ([32, 33]). *Suppose that \mathcal{A} is hypoelliptic on $C^2(\mathcal{G})$ and that Condition 2.1 holds. Then:*

(i) *The solution of (1.1) is globally defined, that is, $\mathbb{P}(\{\tau_\infty = \infty\}) = 1$.*

(ii) *The SDE (1.1) has at least one stationary measure with support contained in \mathcal{G} .*

(iii) *If Condition 2.1 (i) holds, then for any measurable f and almost every sample path $t \mapsto X_t(\omega)$, the limit (1.5) holds, where $\pi \in \mathcal{P}$ has support contained in \mathcal{G} and depends on the starting position of the path, $Z(\omega)$. Furthermore, equation (1.6) holds.*

(iv) If \mathcal{G} is not compact but Condition 2.1 (ii) holds, then $\pi(u) < \infty$ for any $\pi \in \mathcal{P}$ with support contained in \mathcal{G} , and the same as in Theorem 2.2 (iii) is true for every measurable f such that

$$\sup_{x \in \mathcal{G}} \left(\frac{|f(x)|}{1 + |u(x)|} \right) < \infty.$$

Proof. In the case (i), it is easy to argue that the solution is globally defined. In the case (ii), global existence of the solution follows from [33, Thrm.2.1]. The rest follows from Theorems 3.4 and 8.1 in [32], plus Theorem 4.7 in [33] in the case of the drift condition. \square

Remark (The ensemble averages): If, additionally to the premise of Theorem 2.2, the semi-group generated by (1.1) is aperiodic [32], then the analogous statements to Theorem 2.2 (iii) and (iv) hold for the ensemble averages $\mathbb{E}[f(X_t)]$. In this case, we have that

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(X_t)] = \int f(x) \tilde{\pi}(dx),$$

where $\tilde{\pi} \in \mathcal{P}$ depends on the law of the initial condition Z , see [32, Thrm.8.1].

2.3 A relationship between the generator of (1.1) and its stationary measures. The following technical lemma is necessary for the development of the algorithm presented in this paper.

An application of Itô's formula shows that, if $h \in C^2(\mathbb{R}^n)$, then

$$D_t^h := h(X_t) - h(X_0) - \int_0^t \mathcal{A}h(X_s) ds, \quad (2.1)$$

is a local martingale. The generator \mathcal{A} evaluated at function h and point x describes the rate of change in time of the expected value of $h(X_t)$ conditioned on the event $\{X_t = x\}$. If π is the law of X_t , then $\pi(\mathcal{A}h)$ describes the rate of change in time of $\mathbb{E}[h(X_t)]$. It follows that if π is a stationary measure of (1.1), then the law of X_t does not change in time, and thus we would expect that $\pi(\mathcal{A}h) = 0$. Unfortunately, for technical reasons, this is not always the case (see [8] for counterexamples). However, it is not difficult to find sufficient conditions on h such that $\pi(\mathcal{A}h) = 0$. The following lemma gives one such condition specialised for polynomial functions h , which are the focus of this paper. For a proof of the lemma see Appendix A.

Lemma 2.3. *Let π be a stationary measure of (1.1) whose moments of order d exist and are finite. Let h be polynomial with degree $d_h \leq d - \max_{i,j} \{d_{b_i}, d_{a_{ij}}\}$, where b is the drift vector and $a := \sigma\sigma^T$ is the diffusion matrix of (1.1). Then*

$$\pi(\mathcal{A}h) = 0.$$

3 The algorithm. Our algorithm constructs a tractable *outer approximation* $C_{f,\mathcal{G}}^d$ of the set $B_{f,\mathcal{G}}^d$ defined by (1.3). As shown below, the approximation we derive is the image of a *spectrahedron* (a set defined by linear equalities and semidefinite inequalities) through a linear functional. Finding the infimum and supremum of $C_{f,\mathcal{G}}^d$ reduces to solving two SDPs, which can be efficiently carried out using one of several high-quality solvers freely available. Since $B_{f,\mathcal{G}}^d \subseteq C_{f,\mathcal{G}}^d$, the computed infimum (supremum) is a lower (upper) bound of $B_{f,\mathcal{G}}^d$.

To introduce our method, we first take a closer look at the set $B_{f,\mathcal{G}}^d$ from two alternative perspectives. First, note that the set $B_{f,\mathcal{G}}^d$ defined in (1.3) is the image of the set

$$\mathcal{P}_{\mathcal{G}}^d := \{\pi \in \mathcal{P} : \pi \text{ has finite } d\text{-th order moments and support contained in } \mathcal{G}\}$$

through the linear functional (on the vector space of signed measures) $\pi \mapsto \pi(f)$. It is straightforward to check that, as a subset of the vector space of signed measures, $\mathcal{P}_{\mathcal{G}}^d$ is convex. Consequently, its image $B_{f,\mathcal{G}}^d$ is a (possibly unbounded) interval, which is fully described by its supremum and infimum (leaving aside whether or not $B_{f,\mathcal{G}}^d$ contains its endpoints).

Alternatively, since f is a polynomial (with vector of coefficients \mathbf{f}), the set $B_{f,\mathcal{G}}^d$ is the image of the set

$$\mathcal{Y}_{\mathcal{G}}^d := \{y \in \mathbb{R}^{r(d)} : y \text{ is a vector of moments up to order } d \text{ of a measure in } \mathcal{P}_{\mathcal{G}}^d\}$$

through the linear functional (on $\mathbb{R}^{r(d)}$) $y \mapsto \langle y, \mathbf{f} \rangle$. We now see some concrete examples of these sets.

Example 3.1. *To introduce our ideas, we use an example for which there is an extensive body of results. Consider the two-dimensional SDE*

$$dX_t = -\frac{1}{2}X_t dt + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} X_t dW_t, \quad X_0 = Z. \quad (3.1)$$

Applying Itô's formula gives $d\|X_t\| \equiv 0$; hence, $\|X_t\| = \|X_0\|, \forall t \geq 0$. If the initial condition Z takes values in the circle of radius R , \mathbb{S}_R^1 , then the paths of X remain in \mathbb{S}_R^1 . Using Hörmander's condition [16] it is easy to verify that, for each $R > 0$, the generator \mathcal{A} is a hypoelliptic operator on $C^2(\mathbb{S}_R^1)$. By compactness of \mathbb{S}_R^1 , Condition 2.1 (i) is satisfied, and Theorem 2.2 states that, for each $R > 0$, (3.1) has at least one stationary measure with support contained in \mathbb{S}_R^1 . It is also well known [21] that for each R , (3.1) has only one such measure π_R , which is the uniform distribution on \mathbb{S}_R^1 . Therefore, for a given R and f

$$\pi_R(f) = \frac{1}{2\pi} \int_0^{2\pi} f(R \cos(\theta), R \sin(\theta)) d\theta. \quad (3.2)$$

Note that if $R = 0$, then $X \equiv 0$, hence $\pi_0 := \delta_0$, the Dirac measure at zero, is also a stationary measure of (3.1). Consequently, for any d :

$$\mathcal{P}_{\mathbb{S}_R^1}^d = \{\pi_R\}, \quad \mathcal{P}_{\mathbb{R}^2}^d = \{\pi_R : R \geq 0\}, \quad \mathcal{P}_{\mathcal{G}}^d = \emptyset \quad \text{for all other varieties } \mathcal{G}.$$

By the symmetry of the measure π_R , it is easy to show that

$$y_{(\alpha_1, \alpha_2)} := \pi_R(x_1^{\alpha_1} x_2^{\alpha_2}) = \begin{cases} \frac{R^2 \pi^{3/2} 2^{|\alpha|}}{|\alpha|! \Gamma\left(\frac{1-\alpha_1}{2}\right) \Gamma\left(\frac{1-\alpha_2}{2}\right) \Gamma\left(\frac{1-|\alpha|}{2}\right)} & \text{if } \alpha_1 \text{ and } \alpha_2 \text{ are even} \\ 0 & \text{otherwise,} \end{cases}$$

where $\Gamma(\cdot)$ denotes the gamma function. It is now straightforward to describe the sets $\mathcal{Y}_{\mathcal{G}}^d$. For instance, consider the sets containing the $r(2)$ -dimensional vectors of moments up to order $d = 2$ defined as $y := (y_{(0,0)}, y_{(1,0)}, y_{(0,1)}, y_{(2,0)}, y_{(1,1)}, y_{(0,2)}) \in \mathbb{R}^6$. Then we have

$$\mathcal{Y}_{\mathbb{S}_R^1}^2 = \left\{ \left(1, 0, 0, \frac{R^2}{2}, 0, \frac{R^2}{2} \right) \right\} \quad \text{and} \quad \mathcal{Y}_{\mathbb{R}^2}^2 = \{1\} \times \{0\} \times \{0\} \times [0, \infty) \times \{0\} \times [0, \infty).$$

Using the above descriptions of the sets $\mathcal{P}_{\mathcal{G}}^d$ together with the map $\pi \mapsto \pi(f)$ or, alternatively, the sets $\mathcal{Y}_{\mathcal{G}}^d$ together with the map $y \mapsto \langle y, \mathbf{f} \rangle$, we can deduce any projection of interest $B_{f,\mathcal{G}}^d$. For instance,

$$\begin{aligned} \text{for } f_1(x) &= x_1 x_2, & B_{f_1, \mathbb{S}_R^1}^d &= \{0\} & \text{and} & & B_{f_1, \mathbb{R}^2}^d &= \{0\} \\ \text{for } f_2(x) &= x_2^2 + 1, & B_{f_2, \mathbb{S}_R^1}^d &= \left\{ \frac{R^2}{2} + 1 \right\} & \text{and} & & B_{f_2, \mathbb{R}^2}^d &= [1, \infty) \\ \text{for } f_3(x) &= x_2 - 2x_1^2 + 3, & B_{f_3, \mathbb{S}_R^1}^d &= \{3 - R^2\} & \text{and} & & B_{f_3, \mathbb{R}^2}^d &= (-\infty, 3]. \end{aligned}$$

In the above example, we could obtain the sets $\mathcal{P}_{\mathcal{G}}^d$, $\mathcal{Y}_{\mathcal{G}}^d$ and their projections $B_{f,\mathcal{G}}^d$ directly from (1.1). However, this is difficult in general. Indeed, results from real algebraic geometry show that optimising over the cone of vectors whose components are moments of a measure is an NP-hard problem [25]. We believe that, except for trivial cases, the same holds for $\mathcal{Y}_{\mathcal{G}}^d$ which is a subset of this cone. Instead, we construct here a spectrahedral outer approximation $\mathcal{O}_{\mathcal{G}}^d$ of the set $\mathcal{Y}_{\mathcal{G}}^d$. Optimising over $\mathcal{O}_{\mathcal{G}}^d$ consists of solving an SDP, a polynomial-time problem. Explicitly, $\mathcal{O}_{\mathcal{G}}^d$ is a subset of $\mathbb{R}^{r(d)}$ defined by linear equalities and semidefinite inequalities such that $\mathcal{Y}_{\mathcal{G}}^d \subseteq \mathcal{O}_{\mathcal{G}}^d$. Because the outer approximation is a convex set, its image through the linear functional $y \mapsto \langle y, \mathbf{f} \rangle$ is an interval that contains $B_{f,\mathcal{G}}^d$:

$$B_{f,\mathcal{G}}^d \subseteq C_{f,\mathcal{G}}^d := \left\{ \langle y, \mathbf{f} \rangle : y \in \mathcal{O}_{\mathcal{G}}^d \right\}. \quad (3.3)$$

Hence our task is reduced to obtaining the outer approximation $\mathcal{O}_{\mathcal{G}}^d$. We do this in two steps:

- In Section 3.1, we use Lemma 2.3 to construct a set of linear equalities satisfied by the moments of any stationary measure of (1.1).
- In Section 3.2, we construct a set of linear equalities and a semidefinite inequality satisfied by the moments of any unsigned measure with support on \mathcal{G} .

The outer approximation $\mathcal{O}_{\mathcal{G}}^d$ then consists of the set of vectors in $\mathbb{R}^{r(d)}$ that satisfy both of the above.

3.1 Linear equalities satisfied by the moments of stationary measures. By assumption, both the drift vector and the diffusion coefficients in (1.1) are polynomials. Therefore, if h is a polynomial, $\mathcal{A}h$ is also a polynomial. Suppose that y belongs to $\mathcal{Y}_{\mathcal{G}}^d$ and choose any measure π in $\mathcal{P}_{\mathcal{G}}^d$ that has y as its vector of moments of order d . From Lemma 2.3, if $d \geq d_{\mathcal{A}} := \max\{d_{b_i}, d_{a_{ij}}\}$, then

$$\langle y, \mathcal{A}h \rangle = \sum_{\beta \in \mathbb{N}_d^n} (\mathcal{A}h)_{\beta} y_{\beta} = \sum_{\beta \in \mathbb{N}_d^n} (\mathcal{A}h)_{\beta} \pi(x^{\beta}) = \pi \left(\sum_{\beta \in \mathbb{N}_d^n} (\mathcal{A}h)_{\beta} x^{\beta} \right) = \pi(\mathcal{A}h) = 0, \quad \forall h \in \mathbb{R}[x]_{d-d_{\mathcal{A}}}.$$

Since $\{x^{\alpha} : \alpha \in \mathbb{N}_{d-d_{\mathcal{A}}}^n\}$ spans $\mathbb{R}[x]_{d-d_{\mathcal{A}}}$, checking that y satisfies the above is equivalent to checking that

$$\langle y, \mathcal{A}x^{\alpha} \rangle = 0, \quad \forall \alpha \in \mathbb{N}_{d-d_{\mathcal{A}}}^n. \quad (3.4)$$

In words, every vector in $\mathcal{Y}_{\mathcal{G}}^d$ satisfies the $r(d - d_{\mathcal{A}})$ linear equalities defined by (3.4).

At this point, it is worth remarking that the conditions on h and π in Lemma 2.3 are only sufficient but not necessary for $\pi(\mathcal{A}h) = 0$ to hold, as the following example shows.

Example 3.2. Let λ be a positive even integer, and consider the one-dimensional SDE

$$dX_t = (1 - \lambda X_t)dt + \sqrt{2}X_t dW_t, \quad X_0 = Z. \quad (3.5)$$

It is straightforward to verify that Condition 2.1 (ii) holds with $u(x) := x^{\lambda}$ and $\mathcal{G} := \mathbb{R}$, and to use Hörmander's condition to establish that the generator of (3.5) is hypoelliptic on $C^2(\mathbb{R})$. From Theorem 2.2, it follows that: (3.5) has at least one stationary measure; that all of its stationary measures have moments up to order λ ; and that (1.5) holds for any $f \in \mathbb{R}[x]_{\lambda}$. From (3.4), we deduce that given the moments of any such stationary measure $y \in \mathbb{R}^{\lambda+1}$, then

$$k(y_{k-1} - (\lambda + 1 - k)y_k) = 0, \quad k = 1, 2, \dots, \lambda - 2. \quad (3.6)$$

We are only interested in solutions to these equations that are moments of a probability measure. Hence we can append the normalisation $y_0 = 1$ and solve (3.6) to obtain

$$y_k = \prod_{j=0}^{k-1} \frac{1}{\lambda - j}. \quad (3.7)$$

In fact, (3.7) holds for $k = 1, \dots, \lambda$ (instead of only for $k = 1, \dots, \lambda - 2$). This is easily deduced by solving the Fokker-Planck equation associated with (3.5) and showing that the density of a stationary measure of (3.5) is given by the inverse Gamma distribution $\frac{1}{\lambda!} x^{-\lambda-2} e^{-\frac{1}{x}}$. Indeed, the moments of this distribution are given by (3.7) for $k = 1, \dots, \lambda$. Additionally employing the Support Theorem of Stroock and Varadhan [2, Thm.6.6], we can conclude that this is the only stationary measure of (3.5). In conclusion, the moments of any stationary measure of (3.5) satisfy (3.6) for $k = 1, \dots, \lambda$ although Lemma 2.3 only implies that they are satisfied for $k = 1, \dots, \lambda - 2$.

For most SDEs, $y_0 = 1$ together with the equations (3.4) defines a set of underdetermined linear equations as the following example illustrates.

Example 3.3. Consider the SDE

$$dX_t = (1 - 2X_t^3)dt + \sqrt{2}X_t dW_t, \quad X_0 = Z. \quad (3.8)$$

It is straightforward to verify that Condition 2.1 (ii) is satisfied with $u(x) := e^{x^2/2}$ and $\mathcal{G} := \mathbb{R}$, and to use Hörmander's condition to establish that the generator of (3.8) is hypoelliptic on $C^2(\mathbb{R})$. Theorem 2.2 then establishes that (3.8) has globally defined solutions; that it has at least one stationary measure; that all stationary measures have all moments finite; and that (1.5) holds for any $f \in \mathbb{R}[x]$.

Equations (3.4) in this case read:

$$k(y_{k-1} + (k-1)y_k - 2y_{k+2}) = 0, \quad k = 1, 2, \dots$$

By appending $y_0 = 1$ to the above, we can only solve for $y_3 = y_0/2 = 1/2$. The set of equations formed by the first $\tilde{k} \geq 2$ conditions (together with $y_0 = 1$) is underdetermined, and no other moment can be solved for.

3.2 Moment conditions. That the moment equations (3.4) are underdetermined in the above example is essentially the same issue that *moment closure methods* attempt to address (see [39, §3.4] for a review). These methods ‘close’ the equations (3.4) by heuristically removing the dependence of the first few equations on higher moments. We do not follow this approach here. Instead, we overcome this issue by exploiting the fact that we are not interested in *all* the solutions to the equations (3.4), but only in those that are the vector of moments of a probability measure with support contained in \mathcal{G} . There are various tractable conditions, known as *moment conditions*, which are satisfied by the vector of moments of any measure with support contained in \mathcal{G} , but not in general by an arbitrary vector of real numbers. For example, trivially, the even moments of an unsigned Borel measure on \mathbb{R}^n must be non-negative. We now describe in more detail the moment conditions we use in the rest of this paper.

Let π be a measure with vector of moments y of order d , and let g be a polynomial of degree d . If π has support contained in $\{x \in \mathbb{R}^n : g(x) = 0\}$, then

$$\langle y, g \rangle = \sum_{\alpha \in \mathbb{N}_d^n} g_\alpha y_\alpha = \sum_{\alpha \in \mathbb{N}_d^n} g_\alpha \pi(x^\alpha) = \pi\left(\sum_{\alpha \in \mathbb{N}_d^n} g_\alpha x^\alpha\right) = \pi(g) = 0.$$

By the definition (1.4), g_j is zero everywhere in \mathcal{G} for all $j = 1, 2, \dots, \ell$. Thus $\langle y, \mathbf{g}_j \mathbf{h} \rangle = 0$ for every j and $h \in \mathbb{R}[x]_{d-d_{g_j}}$. Since $\{x^\alpha : \alpha \in \mathbb{N}_{d-d_j}^n\}$ spans $\mathbb{R}[x]_{d-d_j}$, checking that y satisfies $\langle y, \mathbf{g}_j \mathbf{h} \rangle = 0$ for any given $h \in \mathbb{R}[x]_{d-d_{g_j}}$ is equivalent to checking that y satisfies the following $r(d-d_{g_j})$ equations:

$$\langle y, \mathbf{g}_j \mathbf{x}^\alpha \rangle = 0, \quad \forall \alpha \in \mathbb{N}_{d-d_{g_j}}^n. \quad (3.9)$$

To these linear equations, we append a semidefinite inequality that stems from the fact that probability measures are unsigned. Let h be any polynomial of degree $s(d) := \lfloor d/2 \rfloor$. Then it follows that

$$(h(x))^2 = \langle m_{s(d)}(x), \mathbf{h} \rangle^2 = \langle \langle m_{s(d)}(x), \mathbf{h} \rangle m_{s(d)}(x), \mathbf{h} \rangle = \langle (m_{s(d)}(x) \otimes m_{s(d)}(x)) \mathbf{h}, \mathbf{h} \rangle.$$

Since h^2 is a non-negative function we have that

$$0 \leq \pi(h^2) = \pi\left(\langle (m_{s(d)} \otimes m_{s(d)}) \mathbf{h}, \mathbf{h} \rangle\right) = \langle \pi(m_{s(d)} \otimes m_{s(d)}) \mathbf{h}, \mathbf{h} \rangle = \langle M_{s(d)}(y) \mathbf{h}, \mathbf{h} \rangle, \quad (3.10)$$

where the *moment matrix*

$$M_{s(d)}(y) := \pi(m_{s(d)} \otimes m_{s(d)}) \in \mathbb{R}^{r(s(d)) \times r(s(d))}$$

denotes the element-wise integration of the matrix $m_{s(d)} \otimes m_{s(d)}$. Note that the entries of the moment matrix are a function of the moments of π :

$$(M_{s(d)}(y))_{\alpha\beta} := y_{\alpha+\beta}, \quad \forall \alpha, \beta \in \mathbb{N}_{s(d)}^n.$$

Since (3.10) holds for all $h \in \mathbb{R}[x]_{s(d)}$, $M_{s(d)}(y)$ is positive semidefinite:

$$M_{s(d)}(y) \succeq 0. \quad (3.11)$$

We then combine (3.4), (3.9), (3.11) and the normalisation $y_0 = 1$ to obtain our outer approximation:

$$\mathcal{O}_{\mathcal{G}}^d := \left\{ y \in \mathbb{R}^{r(d)} : \begin{array}{l} y_0 = 1 \\ \langle y, \mathbf{A} \mathbf{x}^\alpha \rangle = 0, \forall \alpha \in \mathbb{N}_{d-d_A}^n \\ \langle y, \mathbf{g}_j \mathbf{x}^\alpha \rangle = 0, \forall \alpha \in \mathbb{N}_{d-d_{g_j}}^n, \forall j = 1, \dots, \ell \\ M_{s(d)}(y) \succeq 0 \end{array} \right\} \supseteq \mathcal{Y}_{\mathcal{G}}^d. \quad (3.12)$$

From (3.3), it follows that the projection $C_{f,\mathcal{G}}^d = \{\langle y, \mathbf{f} \rangle : y \in \mathcal{O}_{\mathcal{G}}^d\}$ contains $B_{f,\mathcal{G}}^d$. Therefore, we have the following bounds:

$$\rho_{f,\mathcal{G}}^d := \inf C_{f,\mathcal{G}}^d \leq \inf B_{f,\mathcal{G}}^d, \quad \sup B_{f,\mathcal{G}}^d \leq \sup C_{f,\mathcal{G}}^d =: \eta_{f,\mathcal{G}}^d. \quad (3.13)$$

Computing $\rho_{f,\mathcal{G}}^d$ and $\eta_{f,\mathcal{G}}^d$ can be efficiently done by solving two SDPs with $r(d)$ variables each.

Example 3.4. Consider again SDE (3.8) from Example 3.3, whose stationary measures have moments of all orders. Hence

$$B_{x,\mathbb{R}}^1 = B_{x,\mathbb{R}}^2 = \dots = B_{x,\mathbb{R}}^\infty.$$

To find bounds on the mean of the stationary measures of the SDE we construct the outer approximations of $B_{x,\mathbb{R}}^\infty$, as described above. The first few such approximations are:

$$C_{x,\mathbb{R}}^1 = \{y_1 : y \in \mathbb{R}^2, \quad y_0 = 1, \quad y_0 \geq 0\},$$

$$\begin{aligned}
C_{x,\mathbb{R}}^2 &= \left\{ y_1 : y \in \mathbb{R}^3, \quad y_0 = 1, \quad \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} \succeq 0 \right\}, \\
C_{x,\mathbb{R}}^3 &= \left\{ y_1 : y \in \mathbb{R}^4, \quad y_0 = 1, \quad \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} \succeq 0 \right\}, \\
C_{x,\mathbb{R}}^4 &= \left\{ y_1 : y \in \mathbb{R}^5, \quad \begin{array}{l} y_0 = 1 \\ y_0 - y_3 = 0 \end{array} \quad \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0 \right\}, \\
C_{x,\mathbb{R}}^5 &= \left\{ y_1 : y \in \mathbb{R}^6, \quad \begin{array}{l} y_0 = 1 \\ y_0 - y_3 = 0 \\ y_1 + y_2 - y_4 = 0 \end{array} \quad \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0 \right\}, \\
C_{x,\mathbb{R}}^6 &= \left\{ y_1 : y \in \mathbb{R}^7, \quad \begin{array}{l} y_0 = 1 \\ y_0 - y_3 = 0 \\ y_1 + y_2 - y_4 = 0 \\ y_2 + 2y_3 - y_5 = 0 \end{array} \quad \begin{bmatrix} y_0 & y_1 & y_2 & y_3 \\ y_1 & y_2 & y_3 & y_4 \\ y_2 & y_3 & y_4 & y_5 \\ y_3 & y_4 & y_5 & y_6 \end{bmatrix} \succeq 0 \right\}.
\end{aligned}$$

Since a matrix is positive semidefinite if and only if all its principal minors are non-negative, it follows trivially that $C_{x,\mathbb{R}}^1 = C_{x,\mathbb{R}}^2 = C_{x,\mathbb{R}}^3 = \mathbb{R}$. Hence optimising over these sets yields uninformative bounds: $\rho_{x,\mathbb{R}}^1 = \rho_{x,\mathbb{R}}^2 = \rho_{x,\mathbb{R}}^3 = -\infty$ and $\eta_{x,\mathbb{R}}^1 = \eta_{x,\mathbb{R}}^2 = \eta_{x,\mathbb{R}}^3 = \infty$. The higher order approximations $d > 4$, however, lead to non-trivial bounds (Table 1). To obtain the endpoints of the higher order approximations $C_{x,\mathbb{R}}^4, C_{x,\mathbb{R}}^5, \dots$ we used the SDP-solver SDPT3.

d	≤ 4	5	6	7	8	9	10	11	12	13
$\rho_{x,\mathbb{R}}^d$	$-\infty$	0.4133	0.4134	0.6202	0.6202	0.6365	0.6365	0.6376	0.6377	0.6377
$\eta_{x,\mathbb{R}}^d$	∞	0.8283	0.8282	0.6758	0.6757	0.6495	0.6495	0.6494	0.6494	0.6428
d	14	15	16	17	18	19	20	21	22	23
$\rho_{x,\mathbb{R}}^d$	0.6377	0.6377	0.6377	0.6377	0.6377	0.6377	0.6377	0.6377	0.6377	0.6377
$\eta_{x,\mathbb{R}}^d$	0.6428	0.6404	0.6404	0.6402	0.6402	0.6389	0.6389	0.6387	0.6387	0.6384

Table 1: Bounds on the mean of the stationary measures of the SDE (3.8). In total, 40 bounds were computed taking a total CPU time of 10.3 seconds, averaging 0.26 seconds per bound.

For many SDEs, like those in Examples 3.2 and 3.3, all stationary measures have moments of some order d . If \mathcal{G} is compact, $d = \infty$; otherwise, such a d can be found by verifying a drift condition like the one in Condition 2.1. In such cases, $B_{f,\mathcal{G}}^k = B_{f,\mathcal{G}}^d$ for all $d_f \leq k \leq d$. However, as Example 3.4 shows, this does not hold in general for our outer approximations. Instead, we only have that $C_{f,\mathcal{G}}^k \supseteq C_{f,\mathcal{G}}^d$ for all $d_f \leq k \leq d$. Therefore $\{\rho_{f,\mathcal{G}}^k : d_f \leq k \leq d\}$ and $\{\eta_{f,\mathcal{G}}^k : d_f \leq k \leq d\}$ are monotone non-decreasing (resp. non-increasing) sequences of lower (resp. upper) bounds on $B_{f,\mathcal{G}}^d$. In practice, the best bounds are obtained by solving for the infimum/supremum of $C_{f,\mathcal{G}}^k$ for the largest $d_f \leq k \leq d$ that can be handled computationally by the solver.

4 Applications. We now consider three applications of the algorithm. To compute the bounds presented in this section we used the modelling package GloptiPoly 3 [13] to construct the SDPs corresponding to the outer approximations (3.12) and the solver SDPT3 [46] to solve the SDPs. All computations were carried out on a desktop computer with a 3.4 GHz processor and 16GB of memory running Ubuntu 14.04.

4.1 Langevin diffusions, numerical integration, and an inference problem. The Metropolis Adjusted Langevin Algorithm (MALA) [37] is a popular Markov chain Monte Carlo algorithm (MCMC). MALA can be used to estimate integrals with respect to measures of the form

$$\pi_v(dx) := \frac{e^{v(x)}}{Z_v} dx, \quad (4.1)$$

where dx denotes the Lebesgue measure on \mathbb{R}^n , $v : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth confining potential, and Z_v is the normalising constant $Z_v := \int_{\mathbb{R}^n} e^{v(x)} dx$. It is well known that π_v is the unique stationary measure of the *Langevin diffusion*

$$dX_t = \nabla v(X_t) dt + \sqrt{2} dW_t, \quad X_0 = Z. \quad (4.2)$$

The SDE (4.2) has globally defined solutions and, regardless of the initial condition, the limit (1.5) holds with $\pi := \pi_v$ for all π_v -integrable functions f [37]. MALA proceeds by discretising X , adding a Metropolis accept-reject step to preserve stationarity of π_v , and simulating the resulting chain. The time averages of the simulation then converge to the desired average [37].

Since π_v is the unique stationary measure of (4.2), we can use our algorithm to directly compute bounds on $\pi_v(f)$ when both f and v are polynomials, circumventing any discretisation or simulation. We illustrate this idea with the following simple Bayesian inference problem.

Example 4.1. *The scalar noisy time-varying recurrence equation*

$$z_k = p_1 z_{k-1} + p_2 \frac{z_{k-1}}{1 + z_{k-1}^2} + p_3 \cos(1.2(k-1)) + \xi_k, \quad (4.3)$$

is often used to benchmark parameter and state estimation algorithms [9, 3]. For simplicity, we assume that the state $\{z_k : k = 1, \dots, N\}$ is observable and we focus on the problem of estimating the parameters p_1, p_2 , and p_3 . The additive noise $\{\xi_k : k = 1, \dots, N\}$ is typically taken to be an i.i.d. sequence of normally distributed random variables. Since Gaussianity of random variables is not important in our algorithm, we instead choose $\{\xi_k : k = 1, \dots, N\}$ to be an i.i.d. sequence of random variables with bimodal law

$$\mu_\xi(dx) := \pi_{u_\xi}(dx) \quad \text{with} \quad u_\xi(x) = 3x^2 - x^4.$$

where π_{u_ξ} is as in (4.1), see Fig.1a. Choosing parameters $p := (p_1, p_2, p_3) = (0.5, 2, 1)$ and $z_0 = 2$, we use (4.3) to generate N samples $z := \{z_1, z_2, \dots, z_N\}$, see Fig.1b. The inference problem is to use the generated samples z and the initial condition z_0 to estimate the parameters p .

Taking a Bayesian perspective, we first choose a prior distribution μ_0 over the parameters and then we extract information from the posterior distribution $\mu_{p|z}$, see [42]. Our algorithm can be used to this end if the prior μ_0 is of the form (4.1) with a polynomial potential:

$$\mu_0(dx) := \pi_{u_0}(dx) \quad \text{and} \quad u_0 \text{ is a polynomial.}$$

From Bayes' formula, the posterior also takes the form (4.1):

$$\mu_{p|z}(dx) = \pi_v(dx) \quad \text{with} \quad v(p) = \sum_{k=1}^N u_\xi \left(z_k - \left(p_1 z_{k-1} + p_2 \frac{z_{k-1}}{1 + z_{k-1}^2} + p_3 \cos(1.2(k-1)) \right) \right) + u_0(p).$$

We can then use our algorithm to yield lower and upper bounds on the posterior means $\mu_{p|z}(p_1)$, $\mu_{p|z}(p_2)$ and $\mu_{p|z}(p_3)$, and upper bounds on the total variance

$$\text{var}_{p|z} = \mu_{p|z}((p_1 - \mu_{p|z}(p_1))^2) + \mu_{p|z}((p_2 - \mu_{p|z}(p_2))^2) + \mu_{p|z}((p_3 - \mu_{p|z}(p_3))^2).$$

of the posterior distribution $\mu_{p|z}$. For simplicity, we chose our prior μ_0 to be a unit variance zero mean normal distribution, i.e., $u_0(p) := -||p||^2/2$. The results are shown in Fig.1c,d. For $N \geq 15$ samples, we obtain small upper bounds on the total variance (two orders of magnitude smaller than the lower bounds on the posterior means). For this reason, we expect the posterior distribution to resemble a Dirac measure at the vector of posterior means, indicating that the posterior means are appropriate estimators of the parameters (as confirmed in Fig.1c), thus solving our inference problem.

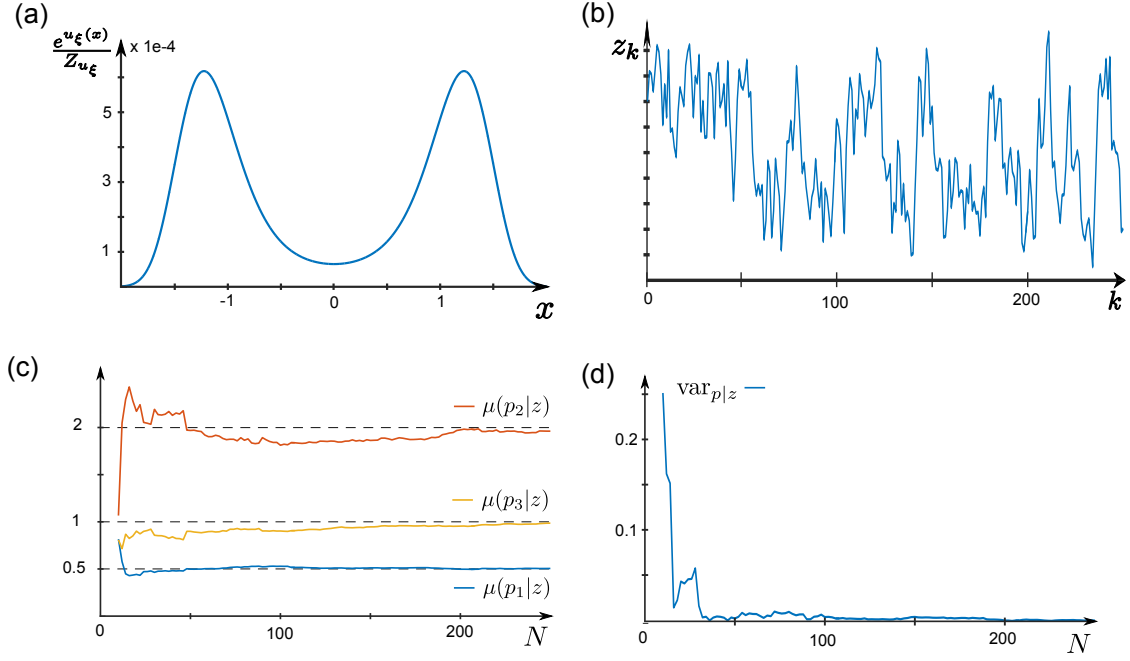


Figure 1: (a) The density of the noise ξ . (b) Sample path of length $N = 250$ generated to estimate the parameters p . The noise was generated by inverting the cdf of μ_ξ numerically (uniform grid on $[-10, 10]$ with step size of 10^{-4}) and drawing independent samples from a uniform $[0, 1]$ distribution. (c) Upper and lower bounds on the means $\mu_{p|z}(p_1)$, $\mu_{p|z}(p_2)$, $\mu_{p|z}(p_3)$ of the posterior distribution $\mu_{z|p}$ (solid lines) plotted against the number of samples $N = 10, 12, \dots, 250$ of z used to generate the posterior. The actual values of the parameters used to generate the samples are shown with dashed lines. The upper and lower bounds were computed by solving (3.13) using $d = 5$ and the appropriate objective $f(x) = x_1, x_2$, or x_3 . The gap between the upper and lower bounds was always smaller than 10^{-2} and hence the upper and lower bounds are indistinguishable in the plot. (d) Upper bounds on the total variance of the posterior distribution, $\text{var}_{p|z} = \mu_{p|z}(p_1^2 + p_2^2 + p_3^2) - \mu_{p|z}(p_1)^2 - \mu_{p|z}(p_2)^2 - \mu_{p|z}(p_3)^2$. These were obtained by computing upper bounds on $\mu_{p|z}(p_1^2 + p_2^2 + p_3^2)$ (solving (3.13) with $d = 5$ and $f(x) = x_1^2 + x_2^2 + x_3^2$), and combining these bounds with the lower bounds computed for the posterior means. In total, 847 bounds were computed taking a total CPU time of 720 seconds, averaging 0.85 seconds per bound.

It is important to remark that Example 4.1 can be solved equally well with MCMC methods—indeed, MCMC algorithms scale better than ours with the dimension of the target measure π_v . However, our alternative approach presents some attractive features: it is fast (see caption of Fig.1) and simple to implement; no tuning of the algorithm is required (e.g., choosing the discretisation step size in MALA); and it delivers deterministic bounds on the integrals of interest instead of

stochastic estimates (hence avoiding issues concerning the convergence of MCMC simulations). Our algorithm can also provide useful information in situations where MCMC methods face difficulties; in particular when the target distribution has several isolated modes. In such cases, MCMC algorithms get stuck at one of the modes and do not explore the rest of the target distribution. Our numerical observations suggest that our algorithm is also affected by the presence of isolated modes, producing a large gap between the upper and lower bounds for the desired integrals (since each bound is stuck at a different mode). The presence of such large gaps can alert the practitioner to the existence of isolated modes, something which is often not obvious for target distributions of dimension three or more. Methods designed to deal with isolated modes, such as simulated-tempering [3], can then be used instead of standard a Metropolis-Hastings MCMC method.

4.2 Lyapunov exponents of linear SDEs driven by multiplicative white noise. In practical applications involving systems of linear ordinary differential equations (ODEs), we are interested in situations where the parameters are perturbed by Gaussian white noise. In those cases, we obtain the following class of linear stochastic differential equations

$$dX(t) = AX(t) dt + \sum_{i=1}^m B_i X(t) dW_i(t), \quad X(0) = Z, \quad (4.4)$$

where $A \in \mathbb{R}^{n \times n}$, $B_i \in \mathbb{R}^{n \times n}$ and $W_i := \{W_i(t) : t \geq 0\}$ are m independent standard Brownian motions on \mathbb{R} . It is well known that (4.4) has globally defined solutions. Furthermore, there exists a jointly continuous process $\{X_t^x : t \geq 0, x \in \mathbb{R}^n\}$ such that $X^Z := \{X_t^Z := X_t \circ Z : t \geq 0\}$ is the unique solution of (4.4) (see [17, Thrm. 21.3]).

In 1967, Khas'minskii [21] made the following observation. Applying Itô's formula twice, he found that the projection of X_t^Z onto the unit sphere, $\Lambda_t^Z := X_t^Z / \|X_t^Z\|$, satisfies the SDE

$$d\Lambda^Z(t) = u_0(\Lambda^Z(t))dt + \sum_{i=1}^m u_i(\Lambda^Z(t))dW_i(t), \quad \Lambda^Z(0) = Z / \|Z\|, \quad (4.5)$$

where $u_0(x) := Ax - \langle x, Ax \rangle x - \sum_{i=1}^m \left(\frac{1}{2} \|B_i x\|^2 x + \langle x, B_i x \rangle B_i x - \frac{3}{2} \langle x, B_i x \rangle^2 x \right)$, and $u_i(x) := B_i x - \langle x, B_i x \rangle x$. In addition, $\Gamma_t^Z := \ln \|X_t^Z\|$ satisfies

$$\Gamma^Z(t) = \ln \|Z\| + \int_0^t Q(\Lambda^Z(s))ds + \sum_{i=1}^m \int_0^t \langle \Lambda^Z(s), B_i \Lambda^Z(s) \rangle dW_i(t), \quad (4.6)$$

where $Q(x) := \langle x, Ax \rangle + \frac{1}{2} \sum_{i=1}^m \|B_i x\|^2 - \langle x, B_i x \rangle^2$. It is not difficult to argue [20, Lem. 6.8] that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \left(\sum_{i=1}^m \int_0^t \langle \Lambda^Z(s), B_i \Lambda^Z(s) \rangle dW_i(t) \right) = 0, \quad \mathbb{P}\text{-almost surely.} \quad (4.7)$$

Suppose that the generator of $\Lambda^Z := \{\Lambda_t^Z : t \geq 0\}$ is hypoelliptic on $C^2(\mathbb{S}^{n-1})$, where $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n : \|x\| = 1\}$ denotes the unit sphere in \mathbb{R}^n . Since Λ^Z , by definition, takes values in \mathbb{S}^{n-1} , Theorem 2.2 tells us that (4.5) has at least one stationary measure, and together with (4.6) and (4.7) that for \mathbb{P} -almost every sample path $t \mapsto X^Z(t)$ there exists a stationary measure π of (4.5) such that

$$\lim_{t \rightarrow \infty} \frac{\Gamma^Z(t)}{t} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t Q(\Lambda^Z(s)) ds = \pi(Q). \quad (4.8)$$

Typically, the above integral is estimated by choosing an appropriate discretisation scheme for (4.4), simulating the resulting chain, and computing the corresponding time average [44]. We instead exploit the fact that u_0, \dots, u_m and Q are all polynomials and apply our algorithm on (4.6) to compute bounds for $\pi(Q)$. In particular, for any $d \geq d_Q$ we have that:

$$\rho_{Q, \mathbb{S}^{n-1}}^d \leq \lim_{t \rightarrow \infty} \frac{\Gamma^Z(t)}{t} \leq \eta_{Q, \mathbb{S}^{n-1}}^d, \quad \mathbb{P}\text{-almost surely,}$$

where $\rho_{Q, \mathbb{S}^{n-1}}^d$ and $\eta_{Q, \mathbb{S}^{n-1}}^d$ are as in (3.13) with notation adapted to (4.5). Note that

$$\begin{aligned} \left\{ \omega \in \Omega : \lim_{t \rightarrow \infty} \frac{\Gamma_t^Z(\omega)}{t} < 0 \right\} &\subseteq \left\{ \omega \in \Omega : \limsup_{t \rightarrow \infty} \|X_t^Z(\omega)\| = 0 \right\}, \\ \left\{ \omega \in \Omega : \lim_{t \rightarrow \infty} \frac{\Gamma_t^Z(\omega)}{t} > 0 \right\} &\subseteq \left\{ \omega \in \Omega : \liminf_{t \rightarrow \infty} \|X_t^Z(\omega)\| = \infty \right\}, \end{aligned}$$

which implies that

$$\begin{aligned} \eta_{Q, \mathbb{S}^{n-1}}^d < 0 &\Rightarrow \mathbb{P} \left(\left\{ \limsup_{t \rightarrow \infty} \|X_t^Z\| = 0 \right\} \right) = 1, \quad \text{for any initial condition } Z, \\ \rho_{Q, \mathbb{S}^{n-1}}^d > 0 &\Rightarrow \mathbb{P} \left(\left\{ \liminf_{t \rightarrow \infty} \|X_t^Z\| = \infty \right\} \right) = 1, \quad \text{for any initial condition } Z, \end{aligned}$$

i.e., the equilibrium solution $X^0 \equiv 0$ of (4.4) is almost surely asymptotically stable if $\eta_{Q, \mathbb{S}^{n-1}}^d < 0$, and almost surely asymptotically unstable if $\rho_{Q, \mathbb{S}^{n-1}}^d > 0$. Therefore, our algorithm applied to (4.5) yields a *sufficient* test for the asymptotic stability or instability of (4.4). The method also yields a *necessary* test for asymptotic stability in the following sense:

Theorem 4.2. *Suppose that the generator of Λ^Z is hypoelliptic on $C^2(\mathbb{S}^{n-1})$. Let*

$$\lambda(Z) := \lim_{t \rightarrow \infty} \frac{\log \|X_t^Z\|}{t}, \quad \lambda_- := \inf_Z \lambda(Z), \quad \lambda_+ := \sup_Z \lambda(Z),$$

where the infimum and supremum are taken over the set of initial conditions Z that are Borel measurable random variables on \mathbb{R}^n . For sufficiently large d ,

$$\rho_{Q, \mathbb{S}^{n-1}}^d > -\infty, \quad \eta_{Q, \mathbb{S}^{n-1}}^d < \infty.$$

Furthermore,

$$\lim_{d \rightarrow \infty} \rho_{Q, \mathbb{S}^{n-1}}^d = \lambda_-, \quad \lim_{d \rightarrow \infty} \eta_{Q, \mathbb{S}^{n-1}}^d = \lambda_+.$$

Proof. Let \mathcal{A} denote the generator of Λ^Z and not of X^Z . The theorem is proved only for the lower bounds $\rho_{Q, \mathbb{S}^{n-1}}^d$. The proof for the upper bounds is identical. The proof proceeds in three steps:

1. We show that the set of limits $\mathcal{S} := \{\lambda(Z) : Z \text{ is a Borel measurable random variable on } \mathbb{R}^n\}$ is the same as the set $B_{Q, \mathbb{S}^{n-1}} := B_{Q, \mathbb{S}^{n-1}}^1 = \dots = B_{Q, \mathbb{S}^{n-1}}^\infty$.
2. We show that the equalities (3.4) fully characterise the stationary measures of (4.5), in the sense that if π is a Borel probability measure with support contained in \mathbb{S}^{n-1} such that

$$\pi(\mathcal{A}x^\alpha) = 0, \quad \forall \alpha \in \mathbb{N}^n, \tag{4.9}$$

then π is a stationary measure of (4.5).

3. We then only have to apply Theorem 4.3 in [25], which shows that, for large enough d , $\rho_{Q, \mathbb{S}^{n-1}}^d > -\infty$, and

$$\lim_{d \rightarrow \infty} \rho_{Q, \mathbb{S}^{n-1}}^d = \inf \left\{ \pi(Q) : \begin{array}{l} \pi \text{ is a Borel probability measure with support contained in } \mathbb{S}^{n-1} \\ \text{that satisfies } \pi(\mathcal{A}x^\alpha) = 0, \forall \alpha \in \mathbb{N}^n \end{array} \right\}.$$

Then (3.4) fully characterises the stationary measures of (4.5), which implies that

$$\lim_{d \rightarrow \infty} \rho_{Q, \mathbb{S}^{n-1}}^d = \inf B_{Q, \mathbb{S}^{n-1}} = \inf \mathcal{S} = \lambda_-.$$

For the detailed proof, see Appendix B. □

Example 4.3. *We exemplify the use of our algorithm in computing Lyapunov exponents through a classic example from [21]. The question of whether an unstable linear system of ODEs could be stabilised by physically realisable multiplicative noise (i.e., multiplicative noise in the sense of Stratonovich) received ample attention in the 1960s. It was shown that this cannot be achieved in one-dimensional systems, and it was hypothesised that it could not be achieved in higher dimensional systems either [30]. Khas'minskii [21] disproved this hypothesis with the following counterexample:*

$$\begin{aligned} dX_1(t) &= c_1 X_1(t) \circ dt + \sigma (X_1(t) \circ dW_1(t) + X_2 \circ dW_2(t)) \\ dX_2(t) &= c_2 X_2(t) \circ dt + \sigma (X_2(t) \circ dW_1(t) - X_1 \circ dW_2(t)) \end{aligned} \quad (4.10)$$

where $c_1 > 0$, $c_2 < 0$, $\sigma > 0$, and $\circ d$ denotes the Stratonovich differential. For this two-dimensional SDE, the projection (4.5) lives in a one dimensional manifold (the unit circle), and thus by changing to polar coordinates one can find analytical expressions for its stationary measures. In this case there is a unique stationary measure and

$$\pi(Q) = \frac{\int_0^{2\pi} \left(\frac{\sigma^2}{2} + c_1 \cos^2(\phi) + c_2 \sin^2(\phi) \right) e^{\frac{(c_1 - c_2)}{\sigma^2} \cos^2(\phi)} d\phi}{\int_0^{2\pi} e^{\frac{(c_1 - c_2)}{\sigma^2} \cos^2(\phi)} d\phi}. \quad (4.11)$$

Khas'minskii then argued that one can always find a sufficiently large (in absolute value) c_2 and σ so that $\pi(Q) < 0$, i.e., so that (4.10) is stable.

Instead, we can verify Khas'minskii's findings for given c_1 and c_2 and σ by solving for the bounds (3.13). As an example, we analysed the system (4.10) with $c_1 = 1$ and $c_2 = -30$. From the bounds presented in Fig.2, we conclude that, for these parameters, (4.10) is stable for $3 \lesssim \sigma \lesssim 3.7$ and unstable otherwise.

In Example 4.3, we could have evaluated (4.11) numerically instead of computing (3.13). However, such analytical expressions for stationary measures of (4.5) are not available in higher dimensions. As explained in [1] (see also [48]): “The direct use of Khas'minskii's method to higher dimensional systems has not met with much success because of the difficulty of studying diffusion processes occurring on surfaces of unit hyperspheres in higher dimensional Euclidean spaces”. Our approach complements Khas'minskii's procedure: it is simple to implement; the number of stationary measures of (4.5) is not a limitation; and the fact that the measures have support on the unit sphere is an advantage instead of a disadvantage, as Theorem 4.2 shows.

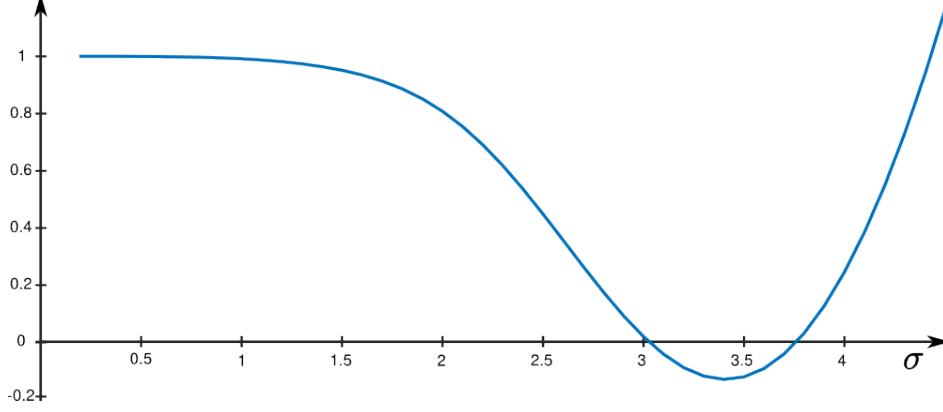


Figure 2: Upper and lower bounds on the Lyapunov exponents of the SDE (4.10) with $c_1 = 1$ and $c_2 = -30$ computed for $\sigma = 0.2, 0.3, \dots, 4.5$ by solving (3.13) with $d = 16$. The gap between the lower and the upper bounds was always smaller than 10^{-3} and hence the two sets of bounds are indistinguishable in the plot. In total 88 bounds were computed for a total CPU time of 364 seconds, averaging 4.1 seconds per bound.

4.3 Piecewise polynomial averages, a noisy nonlinear oscillator and structural reliability problems. Our algorithm can be extended to bound stationary averages $\pi(f)$ where f is a piecewise polynomial of the type

$$f = \sum_{i=1}^N f_i 1_{\mathcal{K}_i}. \quad (4.12)$$

Here f_i are polynomials, 1_A is the indicator function of set A , and \mathcal{K}_i are N disjoint sets in \mathcal{G} defined by

$$\mathcal{K}_i := \{x \in \mathcal{G} : p_j^i(x) = 0, \quad j = 1, \dots, J_i, \quad q_k^i(x) \geq 0, \quad k = 1, \dots, K_i\}. \quad (4.13)$$

This type of extension was first discussed in [24], and has been considered in [25, 27, 7]. Many stationary averages of interest can be written in the above form. For instance, if $f_i := 1, \forall i$, then $\pi(f) = \pi(\cup_i \mathcal{K}_i)$ is the probability that event $\cup_i \mathcal{K}_i$ occurs.

Extension of the algorithm.

We follow here similar arguments to those presented in [25, 27, 7]. Let π be a stationary measure of (1.1) with support in \mathcal{G} and with moments of order d . Let π_i be the restriction of π to set \mathcal{K}_i , and let π_0 denote the restriction to $\mathcal{K}^c := \mathcal{G} \setminus \cup_i \mathcal{K}_i$, i.e., to the complement in \mathcal{G} of the union of all \mathcal{K}_i s. Clearly,

$$\pi = \sum_{i=0}^N \pi_i.$$

Let $y^0, y^1, \dots, y^N \in \mathbb{R}^{r(d)}$ denote the $(N+1)$ vectors of moments of $\pi_0, \pi_1, \dots, \pi_N$, respectively. From the definition (4.12) it follows that

$$\pi(f) = \sum_{i=1}^N \pi_i(f_i 1_{\mathcal{K}_i}) = \sum_{i=1}^N \pi_i(f_i) = \sum_{i=1}^N \langle y^i, \mathbf{f}_i \rangle,$$

assuming that $d_{f_1}, \dots, d_{f_N} \leq d$. The normalisation condition takes the form:

$$\sum_{i=0}^N y_0^i = 1. \quad (4.14)$$

Similarly to (3.4), the proof of Lemma 2.3 tells us that the stationarity of π implies that

$$\left\langle \sum_{i=0}^N y^i, \mathbf{A} \mathbf{x}^\alpha \right\rangle = 0, \quad \forall \alpha \in \mathbb{N}_{d-d_A}^n. \quad (4.15)$$

The vectors of moments $y^i \in \mathbb{R}^{r(d)}$ also fulfil similar conditions to (3.9) and (3.10), in particular:

$$\langle y^i, \mathbf{g}_j \mathbf{x}^\alpha \rangle = 0, \quad \forall \alpha \in \mathbb{N}_{d-d_{g_j}}^n, \quad \forall j = 1, \dots, \ell, \quad \forall i = 0, \dots, N, \quad (4.16)$$

$$M_{s(d)}(y^i) \succeq 0, \quad \forall i = 0, \dots, N. \quad (4.17)$$

Furthermore, the definition (4.13) of the sets \mathcal{K}_i leads to two additional sets of conditions. Firstly, the vectors of moments $y^i \in \mathbb{R}^{r(d)}$ fulfil

$$\langle y^i, \mathbf{p}_j^\dagger \mathbf{x}^\alpha \rangle = 0, \quad \forall \alpha \in \mathbb{N}_{d-d_{p_j}}^n, \quad \forall j = 1, \dots, J_i, \quad \forall i = 1, \dots, N. \quad (4.18)$$

Secondly, using similar arguments as those leading to (3.10), we obtain additional moment conditions. Let q be a polynomial that is non-negative on \mathcal{K}_i and p is any polynomial of degree $s(d-d_q)$. Then we have

$$q(x)(p(x))^2 = \langle (q(x) m_{s(d-d_q)}(x) \otimes m_{s(d-d_q)}(x)) \mathbf{p}, \mathbf{p} \rangle.$$

Since π_i has support in \mathcal{K}_i

$$0 \leq \pi_i(q p^2) = \langle \pi_i(q m_{s(d-d_q)} \otimes m_{s(d-d_q)}) \mathbf{p}, \mathbf{p} \rangle = \langle M_{s(d-d_q)}(\theta_q y^i) \mathbf{p}, \mathbf{p} \rangle \quad (4.19)$$

where $\theta_q : \mathbb{R}^{r(d)} \rightarrow \mathbb{R}^{r(d-d_q)}$ is the shift operator $(\theta_q y)_\alpha = \sum_{\beta \in \mathbb{N}_{d_q}^n} q_\beta y_{\alpha+\beta}$, and the *localising matrix*

$$M_{s(d-d_q)}(\theta_q y^i) := \pi_i(q m_{s(d-d_q)} \otimes m_{s(d-d_q)}) \in \mathbb{R}^{s(d-d_q) \times s(d-d_q)}$$

is the element-wise integration of the matrix $q m_{s(d-d_q)} \otimes m_{s(d-d_q)}$. Since (4.19) holds for all $p \in \mathbb{R}[x]_{s(d-d_q)}$, the localising matrix $M_{s(d-d_q)}(\theta_q y)$ is positive semidefinite. From the definition (4.13), q_k^i is nonnegative on \mathcal{K}_i , hence we obtain our additional set of moment conditions:

$$M_{s(d-d_{q_k^i})}(\theta_{q_k^i} y^i) \succeq 0, \quad \forall k = 1, \dots, K_i, \quad \forall i = 1, \dots, N. \quad (4.20)$$

Together, (4.14)–(4.18) and (4.20) provide computationally tractable necessary conditions satisfied by the moments of any measure with support on \mathcal{K}_i . Thus we have the following outer approximation of $B_{f,\mathcal{G}}^d$:

$$C_{f,\mathcal{G}}^d := \left\{ \sum_{i=1}^N \langle y^i, \mathbf{f}_i \rangle : y^0, \dots, y^N \in \mathbb{R}^{r(d)} \mid (4.14), (4.15), (4.16), (4.17), (4.18), (4.20) \text{ are fulfilled} \right\}. \quad (4.21)$$

The extended algorithm obtains bounds on $B_{f,\mathcal{G}}^d \subseteq C_{f,\mathcal{G}}^d$ by finding the infimum and supremum of $C_{f,\mathcal{G}}^d$; that is, by solving two SDPs with $(N+1)$ times as many variables as in our original algorithm.

Remark (Support of π_0). Notice that $C_{f,\mathcal{G}}^d$ does not include conditions on y^0 related to the support of π_0 . The reason why is that, in general, \mathcal{K}^c has no simple description of the form (4.13). If such a description exists, extra constraints can easily be appended. If such constraints are lacking, there is an unfortunate consequence: $C_{f,\mathcal{G}}^d$ always contains the origin, since $(y^0, y^1, \dots, y^N) := (y, 0, \dots, 0)$ satisfies all the constraints in (4.21). Consequently, our extended algorithm only yields informative bounds if f is nonnegative (or nonpositive) and, in this case, only the upper (resp. lower) bound will be informative. However, for certain purposes this may be sufficient, as the following example demonstrates.

Example 4.4. *We consider the analysis of a noisy nonlinear oscillator in relation to reliability problems of structural mechanics. Structures (e.g., buildings, bridges) perturbed by random forces (e.g. waves, earthquakes) are often modelled with the stationary response of a non-linear oscillator driven by Gaussian white noise [6, 39]. A typical such model used in the literature is the Duffing oscillator*

$$\ddot{Y}_t + \dot{Y}_t + Y_t + \frac{1}{2}Y_t^3 = \sqrt{2}dW_t, \quad (4.22)$$

where $Y_t \in \mathbb{R}$ describes the deviation of the structure from its standing point at time t , and $W := \{W_t : t \geq 0\}$ is a standard Brownian motion. Structural reliability examines whether the structure will bend past a critical value of deviation, and how often this event should be expected to occur [39]. First passage times (i.e., the amount of time it takes for the structure to bend past the critical deviation) and extreme-value probabilities (i.e., how likely the building is to bend past the critical deviation in a given interval of time) are typically investigated. Passage times feature more prominently in the structural mechanics literature since they are more amenable to analysis [6]. Here, we use our extended algorithm to find upper bounds on extreme-value probabilities and on the average fraction of time the structure spends bent beyond the critical deviation.

It is well known [31, §3] that (4.22) has a unique stationary measure

$$\pi(dy \times d\dot{y}) = e^{-\left(\frac{y^2}{2} + \frac{\dot{y}^2}{2} + \frac{y^4}{8}\right)} dy \times d\dot{y}, \quad (4.23)$$

where $dy \times d\dot{y}$ denotes the Lebesgue measure in \mathbb{R}^2 . Throughout this section we assume that the process is at stationarity, i.e., $(Y_t, \dot{Y}_t) \sim \pi, \forall t \geq 0$. We begin by re-writing (4.22) in Itô form

$$dX(t) := \begin{bmatrix} dX_1(t) \\ dX_2(t) \end{bmatrix} = \begin{bmatrix} X_1(t) \\ -X_2(t) - X_1(t) - \frac{1}{2}X_1^3(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ \sqrt{2} \end{bmatrix} dW_t,$$

where $X_1 := Y$ and $X_2 := \dot{Y}$. By Birkhoff's Ergodic Theorem [2, §3], the average fraction of time that the building spends bent beyond the critical deviation u converges to

$$F_u := \pi(1_{\{x \in \mathbb{R}^2 : x_1 \geq u\}}). \quad (4.24)$$

The extreme-value probability is defined as

$$P_u := \mathbb{P} \left(\left\{ \sup_{s \in [0, T]} X_1(s) > u \right\} \right), \quad (4.25)$$

where $[0, T]$ is a given time interval of interest. For sufficiently high u , it is shown in [29] that the up-crossing events become independent; hence, the number of up-crossings in the interval $[0, T]$ has a Poisson distribution with mean $v_u T$, where v_u is the mean threshold crossing rate of u .

For sufficiently regular stationary processes, the mean up-crossing rate can be obtained from Rice's formula [34, 41]

$$v_u := \pi \left(x_2 \mathbf{1}_{\{x \in \mathbb{R}^2: x_1=u, x_2 \geq 0\}} \right), \quad (4.26)$$

whence it follows that

$$P_u \approx 1 - e^{-v_u T}. \quad (4.27)$$

In the computations below, we characterise this regime and consider crossings over high deviations u (at least three times larger than the standard deviation of π); hence we assume that (4.27) holds exactly. From (4.24) and (4.26), it is clear that our extended algorithm can provide upper bounds on v_u and F_u . These bounds are then used with equation (4.27) to bound P_u .

The standard deviation of π is $\sigma \approx 0.761$, as computed directly from (4.23) or using our unmodified algorithm. Following [6], we considered a time interval $T = 100$ and critical deviations u varying from 3σ to 5σ . The largest SDP successfully solved by SDPT3 was $d = 14$, and we computed 14 bounds in a CPU time of 403 seconds, averaging 29 seconds per bound. The upper bounds computed with our algorithm are shown with the exact values computed using (4.23) in Tables 2 and 3. Although the upper bounds are orders of magnitude greater than the true P_u and F_u , the bounds could be useful for practical purposes. For instance, our bounds state that the probability of the structure bending further than five standard deviations at any point over the interval of time $[0, 100]$ is less than 0.011%, and that the structure will spend less than 2 millionth's of a percent of that interval bent beyond this deviation.

This example was chosen so that exact values of F_u and P_u could be computed directly from (4.23), so as to evaluate the quality of the bounds. For most oscillator models, no such analytical expressions are available, and it is often not even clear how many stationary measures exist. Our method applies equally to these other oscillator models.

u/σ	3	$3\frac{1}{3}$	$3\frac{2}{3}$	4	$4\frac{1}{3}$	$4\frac{2}{3}$	5
Upper bound	6.326×10^{-2}	1.673×10^{-2}	4.898×10^{-3}	1.622×10^{-3}	5.905×10^{-4}	2.469×10^{-4}	1.059×10^{-4}
Exact value	4.581×10^{-2}	4.280×10^{-3}	1.974×10^{-4}	4.015×10^{-6}	3.126×10^{-8}	7.978×10^{-11}	5.639×10^{-14}

Table 2: Computed upper bounds on P_u and exact values from (4.27).

u/σ	3	$3\frac{1}{3}$	$3\frac{2}{3}$	4	$4\frac{1}{3}$	$4\frac{2}{3}$	5
Upper bound	4.804×10^{-4}	3.272×10^{-5}	3.781×10^{-6}	8.814×10^{-7}	1.799×10^{-7}	5.903×10^{-8}	1.7954×10^{-8}
Exact value	1.280×10^{-4}	9.267×10^{-6}	3.409×10^{-7}	5.601×10^{-9}	3.560×10^{-9}	7.493×10^{-14}	4.413×10^{-17}

Table 3: Computed upper bounds on F_u and exact values from (4.25).

5 Concluding remarks. In this paper, we have introduced an algorithm based on semidefinite programming that yields upper and lower bounds on stationary averages of SDEs with polynomial drift and diffusion coefficients. The motivation behind our work is the study of long-term behaviour of such SDEs. As explained in the introduction, additional work is required to link the bounds obtained by our algorithm with this long-term behaviour. Typically, a drift condition must be verified by finding an appropriate Lyapunov function [33]. For polynomial drift vectors and diffusion matrices, one can also employ semidefinite programming to search for these Lyapunov functions (see, *sum of squares* programming approaches [35, 36]). In many respects, these approaches are dual to the method we describe in this paper, see [28, 25, 8, 18, 19] for more on the connections.

Our algorithm is also applicable to SDEs whose diffusion coefficients σ are not polynomial, but whose diffusion matrix $a := \sigma\sigma^T$ is polynomial (e.g., the Cox-Ingersoll-Ross interest rate model in financial mathematics). We have concentrated on polynomial diffusion coefficients for simplicity, in

order to guarantee uniqueness of solutions. Furthermore, our algorithm can be extended to SDEs with *rational* drift vector and diffusion matrix—one must just carefully choose polynomials h such that $\mathcal{A}h$ is still a polynomial.

Our choice of moment constraints was motivated by the convenience of use of the modelling package GloptiPoly 3. However, there is a wide selection of moment conditions, some of which lead to easier conic programs (e.g., linear programs or second-order cone programs) [25, 26]. We also restricted ourselves to stationary measures with supports contained in algebraic varieties. We did this to simplify the exposition and because the applications chosen did not require more generality. However, from Section 4.3 it is clear that a similar algorithm can be constructed for measures with supports in so called *basic semialgebraic sets* of the form (4.13) [25]. Such an approach could be advantageous for Example 3.3—using Stroock Varadhan’s Support Theorem it is not difficult to deduce that any stationary measure of those SDEs must have support on the nonnegative semiaxis $[0, \infty)$.

Lastly, a practical issue relating to numerical aspects of our algorithm, and of GMAs in general. In contrast with other approaches, our algorithm runs in polynomial time—its computational complexity follows from that of primal-dual interior point algorithms [23]. However, in our experience, the applicability of the algorithm is affected by certain numerical issues. Specifically, the SDPs that arise from moment problems can be badly conditioned, causing the solvers to perform badly for medium to large problems. We believe that this is a consequence of using raw moments as the basis of the space of moment vectors, as is done in most GMAs. Indeed, the order of magnitude of the moments of a distribution varies rapidly (e.g., see (3.7) in Example 3.2). Since the feasible set $\mathcal{O}_{\mathcal{G}}^d$ over which we optimise contains such a vector of moments and these sets are often compact [25], we expect large discrepancies in the order of magnitude of the entries of our feasible points y and of the moment matrix $M_d(y)$. This can lead to a bad condition number of $M_d(y)$ and consequently to poor performance of the solver. Improvements could be achieved by using an orthonormal basis with respect to the measures of interest, but this is not easy in practice; not only do we usually have little *a priori* information about the measures to guide our choice of basis, but the necessary modifications of the algorithmic implementation are substantial and the package GloptiPoly 3 could not be used in its current form. In our experience, a simple way to mitigate the bad conditioning is to scale the entries of the vectors y by $\tilde{y}_\alpha := y_\alpha / z^\alpha$, where $z \in \mathbb{R}_+^n$. This is equivalent to scaling by the moments of a Dirac measure at z , so z should be chosen so that the entry z_i is close to the absolute value of the i^{th} component of the mean of the measure of interest. A similar scaling was employed in [10]. It is then straightforward to show that \tilde{y} satisfies the same semidefinite inequalities as y , and all that is left to do is to rewrite the equality constraints in terms of the rescaled variables \tilde{y} .

Acknowledgements: We thank Nikolas Kantas and Philipp Thomas for many helpful discussions.

References

- [1] S. T. ARIARATNAM AND W.-C. XIE, *Lyapunov Exponents and Stochastic Stability of Coupled Linear Systems Under Real Noise Excitation*, Journal of Applied Mechanics, 59 (1992), pp. 664–673, doi:10.1115/1.2893775.
- [2] L. R. BELLET, *Ergodic properties of Markov processes*, Open Quantum Systems II, (2006).
- [3] B. P. CARLIN, N. G. POLSON, AND D. S. STOFFER, *A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling*, Journal of the American Statistical Association, 87 (1992), pp. 493–500.
- [4] D. A. DAWSON, *An infinite geostochastic system*, in Multivariate Analysis V, P. R. Krishnaiah, ed., Amsterdam, 1980, North-Holland Publishing Company.
- [5] D. A. DAWSON, *Qualitative behavior of geostochastic systems*, Stochastic Processes and their Applications, 10 (1980), pp. 1–31, doi:10.1016/0304-4149(80)90002-2.

- [6] J. DUNNE AND M. GHANBARI, *Extreme-value prediction for non-linear stochastic oscillators via numerical solutions of the stationary FPK Equation*, Journal of Sound and Vibration, 206 (1997), pp. 697–724, doi:10.1006/jsvi.1997.1148.
- [7] B. ERIKSSON AND M. PISTORIUS, *Method of moments approach to pricing double barrier contracts in polynomial jump-diffusion models*, International Journal of Theoretical and Applied Finance, 14 (2011), pp. 1139–1158, doi:10.1142/S0219024911006644.
- [8] P. W. GLYNN AND A. ZEEVI, *Bounding Stationary Expectations of Markov Processes*, in Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz, E. N. Stewart, J. Feng, and R. H. Stockbridge, eds., vol. 4, Institute of Mathematical Statistics, 2008, pp. 195–214, doi:10.1214/074921708000000381.
- [9] N. J. GORDON, D. J. SALMOND, AND A. F. M. SMITH, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proceedings F - Radar and Signal Processing, 140 (1993), p. 107, doi:10.1049/ip-f-2.1993.0015.
- [10] K. HELMES, S. RÖHL, AND R. H. STOCKBRIDGE, *Computing Moments of the Exit Time Distribution for Markov Processes by Linear Programming*, Operations Research, 49 (2001), pp. 516–530, doi:10.1287/opre.49.4.516.11221.
- [11] K. HELMES AND R. H. STOCKBRIDGE, *Numerical comparison of controls and verification of optimality for stochastic control problems*, Journal of Optimization Theory and Applications, 106 (2000), pp. 107–127, doi:10.1023/A:1004659107996.
- [12] K. HELMES AND R. H. STOCKBRIDGE, *Extension Of Dale’s moment conditions with application to the Wright-Fisher model*, Stochastic Models, 19 (2003), pp. 255–267, doi:10.1081/STM-120020389.
- [13] D. HENRION, J.-B. LASSERRE, AND J. LÖFBERG, *GloptiPoly 3: moments, optimization and semidefinite programming*, Optimization Methods and Software, 24 (2009), pp. 761–779, doi:10.1080/10556780802699201.
- [14] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer New York, New York, NY, 1999, doi:10.1007/978-1-4612-0561-6.
- [15] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Markov Chains and Invariant Probabilities*, Birkhäuser Basel, Basel, 2003, doi:10.1007/978-3-0348-8024-4.
- [16] L. HÖRMANDER, *Hypoelliptic second order differential equations*, Acta Mathematica, 119 (1967), pp. 147–171, doi:10.1007/BF02392081.
- [17] O. KALLENBERG, *Foundations of Modern Probability*, Springer-Verlag, New York Berlin Heidelberg, second ed., 2001.
- [18] K. KASHIMA AND R. KAWAI, *An optimization approach to weak approximation of stochastic differential equations with jumps*, Applied Numerical Mathematics, 61 (2011), pp. 641–650, doi:10.1016/j.apnum.2010.10.012.
- [19] K. KASHIMA AND R. KAWAI, *On Weak Approximation of Stochastic Differential Equations through Hard Bounds by Mathematical Programming*, SIAM Journal on Scientific Computing, 35 (2013), pp. A1–A21, doi:10.1137/110841497.
- [20] R. KHASHINSKII, *Stochastic Stability of Differential Equations*, vol. 66 of Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, Berlin, Heidelberg, second ed., 2012, doi:10.1007/978-3-642-23280-0.
- [21] R. Z. KHASHINSKII, *Necessary and Sufficient Conditions for the Asymptotic Stability of Linear Stochastic Systems*, Theory of Probability & Its Applications, 12 (1967), pp. 144–147, doi:10.1137/1112019.
- [22] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin Heidelberg New York, 4th ed., 2011.
- [23] Y. LABIT, D. PEACELLE, AND D. HENRION, *SEDUMI INTERFACE 1.02: a tool for solving LMI problems with SEDUMI*, in Proceedings. IEEE International Symposium on Computer Aided Control System Design, IEEE, 2002, pp. 272–277, doi:10.1109/CACSD.2002.1036966.
- [24] J.-B. LASSERRE, *Bounds on measures satisfying moment conditions*, Annals of Applied Probability, 12 (2002), pp. 1114–1137, doi:10.1214/aoap/1031863183.
- [25] J. B. LASSERRE, *Moments, Positive Polynomials and Their Applications*, Imperial College Press Optimization Series, Imperial College Press, London, 2009, doi:10.1142/p665.
- [26] J.-B. LASSERRE AND T. PRIETO-RUMEAU, *SDP vs. LP relaxations for the moment approach in some performance evaluation problems*, Stochastic Models, 20 (2004), pp. 439–456, doi:10.1081/STM-200033112.
- [27] J.-B. LASSERRE, T. PRIETO-RUMEAU, AND M. ZERVOS, *Pricing a class of exotic options via moments and SDP relaxations*, Mathematical Finance, 16 (2006), pp. 469–494, doi:10.1111/j.1467-9965.2006.00279.x.

- [28] M. LAURENT, *Sums of Squares, Moment Matrices and Optimization Over Polynomials*, in Emerging applications of algebraic geometry, M. Putinar and S. Sullivant, eds., 2009, pp. 157–270, doi:10.1007/978-0-387-09686-5_7.
- [29] M. R. LEADBETTER, G. LINDGREN, AND H. ROOTZÉN, *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer New York, New York, NY, 1983, doi:10.1007/978-1-4612-5449-2.
- [30] M. LEIBOWITZ, *Statistical Behavior of Linear Systems with Randomly Varying Parameters*, Journal of Mathematical Physics, 4 (1963), p. 852, doi:10.1063/1.1724328.
- [31] J. MATTINGLY, A. STUART, AND D. HIGHAM, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic Processes and their Applications, 101 (2002), pp. 185–232, doi:10.1016/S0304-4149(02)00150-3.
- [32] S. P. MEYN AND R. L. TWEEDIE, *Stability of Markovian Processes II: Continuous-Time Processes and Sampled Chains*, Advances in Applied Probability, 25 (1993), pp. 487–517, doi:10.2307/1427521.
- [33] S. P. MEYN AND R. L. TWEEDIE, *Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes*, Advances in Applied Probability, 25 (1993), pp. 518–548, doi:10.2307/1427522.
- [34] A. NAESS, *Technical Note: On a rational approach to extreme value analysis*, Applied Ocean Research, 6 (1984), pp. 173–174, doi:10.1016/0141-1187(84)90007-5.
- [35] P. A. PARRILO, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, PhD thesis, California Institute for Technology, 2000.
- [36] P. A. PARRILO, *Semidefinite programming relaxations for semialgebraic problems*, Mathematical Programming, 96 (2003), pp. 293–320, doi:10.1007/s10107-003-0387-5.
- [37] G. O. ROBERTS AND R. L. TWEEDIE, *Exponential Convergence of Langevin Distributions and Their Discrete Approximations*, Bernoulli, 2 (1996), pp. 341–363, doi:10.2307/3318418.
- [38] L. C. G. ROGERS AND D. WILLIAMS, *Diffusions, Markov Processes and Martingales: Volume 2. Itô Calculus*, Cambridge University Press, Cambridge, second ed., jun 2000.
- [39] G. I. SCHÜLLER, *A state-of-the-art report on computational stochastic mechanics*, Probabilistic Engineering Mechanics, 12 (1997), pp. 197–321, doi:10.1016/S0266-8920(97)00003-9.
- [40] E. SCHWERER, *A linear programming approach to the steady-state analysis of Markov processes*, PhD Thesis, tech. report, Graduate School of Business, Stanford University, 1996.
- [41] T. SOONG, *Random differential equations in science and engineering*, Academic Press, 1973.
- [42] A. M. STUART, *Inverse problems: A Bayesian perspective*, Acta Numerica, 19 (2010), pp. 451–559, doi:10.1017/S0962492910000061.
- [43] D. TALAY, *Second order discretization schemes of stochastic differential systems for the computation of the invariant law*, Stochastics and Stochastic Reports, 29 (1990), pp. 13–36.
- [44] D. TALAY, *Approximation of upper Lyapunov exponents of bilinear stochastic differential systems*, SIAM Journal on Numerical Analysis, 28 (1991), pp. 1141–1164.
- [45] D. TALAY, *Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Processes and Related Fields, 8 (2002), pp. 163–198.
- [46] K.-C. TOH, M. TODD, AND R. TÜTÜNCÜ, *SDPT3 A Matlab software package for semidefinite programming, Version 1.3*, Optimization Methods and Software, 11 (1999), pp. 545–581.
- [47] S. WATANABE AND N. IKEDA, *Stochastic differential equations and diffusion processes*, Kodansha Scientific Books, Tokyo, 2nd editio ed., 1989.
- [48] Z. WEIQIU AND C. GUOQIANG, *Nonlinear stochastic dynamics: A survey of recent developments*, Acta Mechanica Sinica, 18 (2002), pp. 551–566, doi:10.1007/BF02487958.

A Proof of Lemma 2.3

Lemma 2.3. Let X be a solution of (1.1) whose initial condition has law π . Suppose that D^h (defined in (2.1)) is a martingale and not just a local martingale. Then

$$\mathbb{E} \left[D_t^h \right] = \mathbb{E} \left[\mathbb{E} \left[D_t^h \middle| \mathcal{F}_0 \right] \right] = \mathbb{E} \left[D_0^h \right] = 0,$$

for any $t > 0$. Since π is a stationary measure of (1.1), $\mathbb{E}[h(X_t)] = \mathbb{E}[h(X_0)] = \pi(h)$ and so the above implies

$$\mathbb{E} \left[\int_0^t \mathcal{A}h(X_s) ds \right] = 0.$$

Using Tonelli's Theorem and stationarity we have

$$\mathbb{E} \left[\int_0^t |\mathcal{A}h(X_s)| ds \right] = \int_0^t \mathbb{E} [|\mathcal{A}h(X_s)|] ds = \int_0^t \pi(|\mathcal{A}h|) ds = t\pi(|\mathcal{A}h|).$$

Since h is of degree less or equal than $d - \max\{d_{b_i}, d_{a_{ij}}\}$, our assumptions on π imply that $\pi(|\mathcal{A}h|) < \infty$. Thus, we have that $\mathcal{A}h(X_s)$ is integrable with respect to $\mathbb{P} \times \lambda_t$, where λ_t denotes the Lebesgue measure on $[0, t]$. Choosing any $t > 0$ and applying Fubini's Theorem we obtain

$$t\pi(\mathcal{A}h) = \int_0^t \mathbb{E} [\mathcal{A}h(X_s)] ds = \mathbb{E} \left[\int_0^t \mathcal{A}h(X_s) ds \right] = 0$$

from which it follows that $\pi(\mathcal{A}h) = 0$.

We now need to argue that D^h is indeed a martingale. To show that D^h is a martingale it suffices to show that $\mathbb{E}[D_t^h | \mathcal{F}_s] = D_s^h$ for all $t \geq s$. Equivalently that $\mathbb{E}[1_A D_t^h] = \mathbb{E}[1_A D_s^h]$ for any $A \in \mathcal{F}_s$ and $t \geq s$. We do this by finding a sequence of martingales $\{D^{h_m} : m \in \mathbb{Z}\}$ such that for every $t \geq 0$, $\{D_t^{h_m} : m \in \mathbb{Z}\}$ is dominated by a \mathbb{P} -integrable random variable and such that $\{D_t^{h_m} : m \in \mathbb{Z}\}$ converges almost surely to D_t^h . With such a sequence at hand, we can use dominated convergence and the martingale property of D^{h_m} to establish the desired result:

$$\mathbb{E}[1_A D_t^h] = \lim_{m \rightarrow \infty} \mathbb{E}[1_A D_t^{h_m}] = \lim_{m \rightarrow \infty} \mathbb{E}[1_A D_s^{h_m}] = \mathbb{E}[1_A D_s^h].$$

Thus, all that remains is to construct the sequence $\{D^{h_m} : m \in \mathbb{Z}\}$. We do so by using the fact that if $g \in C^2(\mathbb{R}^n)$ is compactly supported, then D^g is a martingale, see [38, §V]. For any natural number m let

$$h_m(x) := \phi(x_1/m)\phi(x_2/m)\dots\phi(x_n/m)h(x)$$

where ϕ is the smooth compactly supported function

$$\phi(y) := \begin{cases} \exp\left(-\frac{y^2}{1-y^2}\right) & \text{if } |y| < 1 \\ 0 & \text{otherwise} \end{cases}.$$

By definition of h_m , we have that $D_t^{h_m}$ tends almost surely to D_t^h and so we just need to show that $D_t^{h_m}$ is dominated by a \mathbb{P} -integrable random variable.

Since $\partial_z \phi(z/m)|_{z=y} = \partial_x \phi(x)/m|_{x=y/m}$, and $\partial_{zz} \phi(z/m)|_{z=y} = \partial_{xx} \phi(x)/m^2|_{x=y/m}$, and $\phi, \partial\phi$, and $\partial^2\phi$ are all bounded (since they are continuous functions non-zero only on the compact set $[-1, 1]$) we have that

$$|\mathcal{A}h_m| \leq \beta \left(\sum_{i=1}^n (|hb_i| + |\partial_i hb_i|) + \sum_{i,j=1}^n |ha_{ij}| + |\partial_i ha_{ij}| + |\partial_j ha_{ij}| + |\partial_{ij} ha_{ij}| \right)$$

where β is a constant that depends on the maximums of $\phi, \partial\phi$, and $\partial^2\phi$. Since all the polynomials in the right-hand side are of degree d or less, the right-hand side is π -integrable. Thus, the sequence of random variables $\{D_t^{h_m}\}_{m \in \mathbb{Z}_+}$ is dominated by

$$|h(X_t)| + |h(X_0)| + \beta \int_0^t \sum_{i=1}^n (|h(X_s)b_i(X_s)| + |\partial_i h(X_s)b_i(X_s)|) + \sum_{i,j=1}^n |h(X_s)a_{ij}(X_s)| + |\partial_i h(X_s)a_{ij}(X_s)|$$

$$+|\partial_j h(X_s)a_{ij}(X_s)| + |\partial_{ij} h(X_s)a_{ij}(X_s)|ds.$$

Using Tonelli's Theorem, and stationarity of X as we did before, we have that the above is \mathbb{P} -integrable and the lemma follows. \square

B Proof of Theorem 4.2

Theorem 4.2. As explained in the main text, all we have to do is carry out Steps 1 and 2.

Step 1: Equation (4.8) tells us that $\mathcal{S} \subseteq B_{Q, \mathbb{S}^{n-1}}$. Conversely, suppose that π is a stationary measure of (4.5) with support contained in \mathbb{S}^{n-1} and choose an initial condition Z with law π . Since $\Lambda^Z(0) = Z/\|Z\| = Z$, stationarity of π and the Markov property of Λ^Z implies that Λ^Z is a stationary process with one-dimensional law π . Thus

$$\lambda(Z) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t Q(\Lambda^Z(s))ds = \pi(Q), \quad \mathbb{P}\text{-almost surely,}$$

where the first equality follows from (4.6) and (4.7), and the second is a consequence of Birkhoff's Ergodic Theorem (for instance, see [2, §3]). So $B_{Q, \mathbb{S}^{n-1}} \subseteq \mathcal{S}$ and we have the desired $\mathcal{S} = B_{Q, \mathbb{S}^{n-1}}$.

Step 2: Suppose that π is a Borel probability measure with support contained in \mathbb{S}^{n-1} that satisfies (4.9). We have to argue that π is a stationary measure of (4.5), that is that if Z has law π , then $\mathbb{P}(\Lambda_t^Z \in A) = \pi(A)$ for each $t \geq 0$ and each Borel measurable set $A \subseteq \mathbb{S}^{n-1}$. By approximating indicator functions with smooth functions, it is enough to argue that $\mathbb{E}[h(\Lambda_t^Z)] = \pi(h)$ for each $t \geq 0$ and each smooth function $h : \mathbb{R}^n \rightarrow \mathbb{R}$.

We first show that $\pi(\mathcal{A}h) = 0$ for all smooth functions h . By linearity, (4.9) implies that $\pi(\mathcal{A}p) = 0$ for all polynomials p . Using Weierstrass' Approximation Theorem it is straightforward to argue that for any smooth h and $\varepsilon > 0$ there exists a polynomial p such that

$$\max_{x \in \mathbb{S}^{n-1}} \left(|h(x) - p(x)| + \sum_{i=1}^n |\partial_i h(x) - \partial_i p(x)| + \sum_{i,j=1}^n |\partial_{ij} h(x) - \partial_{ij} p(x)| \right) \leq \varepsilon. \quad (\text{B.1})$$

see Lemma B.1 below. Using the above we have that

$$|\mathcal{A}h(x) - \mathcal{A}p(x)| \leq \varepsilon \left(\max_{x \in \mathbb{S}^{n-1}} \left(\sum_{i=1}^n |b_i(x)| + \sum_{i,j=1}^n |a_{ij}(x)| \right) \right), \quad \forall x \in \mathbb{S}^{n-1}.$$

Since $\pi(\mathcal{A}p) = 0$, π has finite mass, and ε was arbitrary, it follows that $\pi(\mathcal{A}h) = 0$.

Let us return to arguing that $\mathbb{E}[h(\Lambda_t^Z)] = \pi(h)$, where Z is any initial condition with law π . Since the drift vector and diffusion coefficients of (4.4) are linear, we can find a modification of $\{X_t^x : t \geq 0, x \in \mathbb{R}^n\}$ such that for every $t \geq 0$, the map $\mathbb{R}^n \ni x \mapsto X_t^x \in \mathbb{R}^n$ is smooth, \mathbb{P} -almost surely, see Proposition 2.2 in [47, §5]. Consequently, for every $t \geq 0$, the map $x \mapsto \Lambda_t^x$ is also smooth, \mathbb{P} -almost surely. Now choose any smooth function $h : \mathbb{R}^n \rightarrow \mathbb{R}$. Itô's formula tells us that

$$h(\Lambda_t^Z) = h(\Lambda_0^Z) + \int_0^t (\mathcal{A}h)(\Lambda_s^Z)ds + \int_0^t \langle U(\Lambda_s^Z) \nabla h(\Lambda_s^Z), dW_s \rangle, \quad \mathbb{P}\text{-almost surely,}$$

where U is the matrix with columns u_1, \dots, u_m . By definition, the paths of Λ^Z are contained in a compact set, thus applying the Dominated Convergence Theorem we have that

$$M_t := \int_0^t \langle U(\Lambda_s^Z) \nabla h(\Lambda_s^Z), dW_s \rangle \quad (\text{B.2})$$

is not just a local martingale, but a martingale. Thus

$$\mathbb{E} [h(\Lambda_t^Z)] = \pi(h) + \mathbb{E} \left[\int_0^t (\mathcal{A}h)(\Lambda_s^Z) ds \right].$$

Because for any $x \in \mathbb{S}^{n-1}$ the paths of Λ^x are contained in \mathbb{S}^{n-1} , and because both the drift vector and the diffusion matrix are continuous functions, the map $[0, \infty) \times \mathbb{S}^{n-1} \ni (t, x) \mapsto (\mathcal{A}h)(\Lambda_t^x) \in \mathbb{R}$ is bounded. Thus we can apply Fubini's Theorem to obtain that

$$\mathbb{E} \left[\int_0^t (\mathcal{A}h)(\Lambda_s^Z) ds \right] = \int_0^t \mathbb{E} [(\mathcal{A}h)(\Lambda_s^Z)] ds.$$

So it is clearly enough to argue that $\mathbb{E} [(\mathcal{A}h)(\Lambda_t^Z)] = 0$ for each $t \geq 0$. Deconvolving the expectation we have that

$$\mathbb{E} [(\mathcal{A}h)(\Lambda_t^Z)] = \int \mathbb{E} [(\mathcal{A}h)(\Lambda_t^x)] \pi(dx).$$

Let $u_t(x) := \mathbb{E} [h(\Lambda_t^x)]$. Since $\mathbb{S}^{n-1} \ni x \mapsto h(\Lambda_t^x) \in \mathbb{R}$ is bounded, we can differentiate under the expectation sign. Thus, for each $t \geq 0$, smoothness of h and of $x \mapsto \Lambda_t^x$ implies that $x \mapsto u_t(x)$ is a smooth function too. If we can show that $\mathbb{E} [(\mathcal{A}h)(\Lambda_t^x)] = (\mathcal{A}u_t)(x)$ for each $t \geq 0$, then we are done since $\pi(\mathcal{A}u_t) = 0$. Arguing this fact is routine, see Lemma B.2 below. \square

We find it convenient in the proof of the following lemma to write $\partial_\alpha f$ as a shorthand for

$$\frac{\partial^{\alpha_1} \partial^{\alpha_2} \dots \partial^{\alpha_n} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_n^{\alpha_n}}$$

where α is any tuple in \mathbb{N}^n .

Lemma B.1. *Suppose that \mathcal{M} is a compact smooth embedded submanifold of \mathbb{R}^n , that $h : \mathcal{G} \rightarrow \mathbb{R}$ is smooth and choose $\varepsilon > 0$. Then, there exists a polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ such that (B.1) holds (with \mathbb{S}^{n-1} replaced by \mathcal{M}).*

Proof. Let \tilde{h} be any compactly supported smooth extension of h on \mathbb{R}^n . Choose l such that the support of \tilde{h} is contained in the hypercube $[-l, l]^n$. For any real-valued continuous function f on \mathbb{R}^n , define

$$G_i f(x) := \int_{-l}^{x_i} f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) dy, \quad i = 1, \dots, n$$

and

$$G_\alpha f(x) := \underbrace{G_1 G_1 \dots G_1}_{\alpha_1 \text{ times}} \underbrace{G_2 G_2 \dots G_2}_{\alpha_2 \text{ times}} \dots \underbrace{G_n G_n \dots G_n}_{\alpha_n \text{ times}} f(x), \quad \alpha \in \mathbb{N}^n.$$

Now choose any $\varepsilon > 0$. By the Weierstrass Approximation Theorem there exists a polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$\sup_{x \in \mathcal{M}} |\partial_{\mathbf{2}} \tilde{h}(x) - p(x)| \leq \varepsilon,$$

where $\mathbf{2} := (2, 2, \dots, 2)$ the tuple in \mathbb{N}^n whose entries are all 2. Let $q(x) := G_{\mathbf{2}} p(x)$ and choose any $\alpha \in \mathbb{N}_2^n$. Since $\partial_\alpha \tilde{h}$ has support in $[-l, l]^n$, applying repeatedly the fundamental theorem of calculus we have that $\partial_\alpha \tilde{h} = G_{\mathbf{2}-\alpha} \partial_{\mathbf{2}} \tilde{h}$. Applying repeatedly Leibniz integral rule we then have that

$$\partial_\alpha \tilde{h} - \partial_\alpha q = G_{\mathbf{2}-\alpha} \partial_{\mathbf{2}-\alpha} \partial_\alpha \tilde{h} - \partial_\alpha G_{\mathbf{2}} p = G_{\mathbf{2}-\alpha} (\partial_{\mathbf{2}} \tilde{h} - p)$$

Since for any continuous function f and $i = 1, \dots, n$, it is the case that $|G_i f| \leq G_i |f|$, we have that

$$\begin{aligned} \sup_{x \in \mathcal{M}} |\partial_\alpha h(x) - \partial_\alpha q(x)| &= \sup_{x \in \mathcal{M}} |\partial_\alpha \tilde{h}(x) - \partial_\alpha q(x)| \leq \sup_{x \in \mathcal{M}} \left(G_{2-\alpha} |\partial_2 \tilde{h} - p|(x) \right) \\ &\leq G_{2-\alpha} \left(\sup_{x \in \mathcal{M}} |\partial_2 \tilde{h}(x) - p(x)| \right). \end{aligned}$$

From the definition of p we have that the righthand side of the above is less or equal than $(2l)^{2-\alpha}\varepsilon$. Since the ε was arbitrary, (B.1) follows. \square

Lemma B.2. *Under the conditions of Theorem 4.2 we have that, for every $t \geq 0$ and $x \in \mathbb{R}^n$*

$$\mathbb{E}[(\mathcal{A}h)(\Lambda_t^x)] = (\mathcal{A}u_t)(x)$$

Proof. Choose any $t, s \geq 0$ and $x \in \mathbb{R}^n$. For typographical condition we write we use Λ_t^x and $\Lambda(t, x)$ interchangeably in this proof. If μ_s^x denotes the law of Λ_s^x , by definition of u_t we have that

$$\mathbb{E}[u_t(\Lambda_s^x)] = \int \mathbb{E}[h(\Lambda_t^y)] \mu_s^x(dy) = \mathbb{E}[h(\Lambda(t, \Lambda(s, x)))].$$

By the Markov property of Λ we have that

$$\mathbb{E}[h(\Lambda(t, \Lambda(s, x)))] = \mathbb{E}[h(\Lambda(t+s, x))].$$

Applying Itô's rule we have that

$$h(\Lambda(t+s, x)) = h(\Lambda(t, x)) + \int_t^{t+s} (\mathcal{A}h)(\Lambda_v^x) dv + \int_t^{t+s} \langle U(\Lambda_v^Z) \nabla h(\Lambda_v^Z), dW_v \rangle.$$

Since the paths of Λ take values in a compact set the rightmost term above is a martingale (as a function of s). Taking expectations and applying Fubini's Theorem we have that

$$\mathbb{E}[u_t(\Lambda_s^x)] = u_t(x) + \int_t^{t+s} \mathbb{E}[(\mathcal{A}h)(\Lambda_v^x)] dv$$

Thus,

$$\lim_{s \rightarrow 0} \frac{\mathbb{E}[u_t(\Lambda_s^x)] - u_t(x)}{s} = \mathbb{E}[(\mathcal{A}h)(\Lambda_t^x)]. \quad (\text{B.3})$$

Next, applying Itô's rule we have that

$$u_t(\Lambda_s^x) = u_t(x) + \int_0^s (\mathcal{A}u_t)(\Lambda_v^x) dv + \int_0^s \langle U(\Lambda_v^Z) \nabla u_t(\Lambda_v^Z), dW_v \rangle.$$

Similarly, since the paths of Λ take values in a compact set, the rightmost term is a martingale. Thus taking expectations of the above we have that

$$\mathbb{E}[u_t(\Lambda_s^x)] - u_t(x) = \mathbb{E} \left[\int_0^s (\mathcal{A}u_t)(\Lambda_v^x) dv \right] = s(\mathcal{A}u_t)(x) + \mathbb{E} \left[\int_0^s (\mathcal{A}u_t)(\Lambda_v^x) - (\mathcal{A}u_t)(\Lambda_0^x) dv \right].$$

Applying Itô's rule, taking expectations, exploiting compactness of the paths of Λ , and applying Fubini's Theorem we have that

$$\left| \mathbb{E} \left[\int_0^s (\mathcal{A}u_t)(\Lambda_v^x) - (\mathcal{A}u_t)(\Lambda_0^x) dv \right] \right| = \left| \mathbb{E} \left[\int_0^s \int_0^v \mathcal{A}^2 u_t(\Lambda_w^x) dw \right] \right| \leq s^2 \left(\max_{x \in \mathbb{S}^{n-1}} |(\mathcal{A}^2 u_t)(x)| \right).$$

Consequently,

$$\lim_{s \rightarrow 0} \frac{\mathbb{E}[u_t(\Lambda_s^x)] - u_t(x)}{s} = \mathcal{A}u_t(x).$$

Comparing the above with (B.3) gives the desired result. \square